

UNIVERSITÉ KASDI MERBAH – OUARGLA
FACULTÉ DES HYDROCARBURES, DES ÉNERGIES RENOUVELABLES ET DES
SCIENCES DE LA TERRE ET DE L'UNIVERS
DÉPARTEMENT DES SCIENCES DE LA TERRE ET DE
L'UNIVERS



Doctorat en Sciences

Domaine : Sciences de la Terre et de l'Univers

Filière : Géologie

THEME

**MODÉLISATION GÉOLOGIQUE DES RÉSERVOIRS PÉTROLIERS A L'AIDE DES
MÉTHODES DE L'INTELLIGENCE ARTIFICIELLE, CAS DU BASSIN DE BERKINE,
SUD-EST DE L'ALGÉRIE**

Présenté par

AMEUR-ZAIMECHE Ouafi

Soutenu publiquement le 17/12/2020

Devant le jury :

Président :	Belksier Mohammed Salah	M.C.A	Univ. Ouargla
Directeur de thèse:	Zeddouri Aziez	Professeur	Univ. Ouargla
Co-directeur de thèse :	Heddami Salim	Professeur	Univ. Skikda
Examineur :	Ali Zerrouki Ahmed	M.C.A	Univ. Ouargla
Examineur :	Hadjaj Ahmed	Professeur	Univ. Boumerdès
Examineur :	Chettih Mohammed	Professeur	Univ. Laghouat

Année Universitaire : 2019/2020

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

(وَتَرَى الْجِبَالَ تَحْسَبُهَا جَامِدَةً وَهِيَ تَمُرُّ مَرَّ السَّحَابِ صُنْعَ اللَّهِ الَّذِي أَتَقَنَ كُلَّ شَيْءٍ إِنَّهُ خَبِيرٌ بِمَا تَفْعَلُونَ) (سورة النمل الآية 88)

(Et tu verras les montagnes - tu les crois figées - alors qu'elles passent comme des nuages. Telle est l'œuvre d'Allah qui a tout façonné à la perfection. Il est Parfaitement Connaisseur de ce que vous faites!) (Sourate: An Naml_verse :88)

(And you see the mountains, thinking them rigid, while they will pass as the passing of clouds. [It is] the work of Allah , who perfected all things. Indeed, He is Acquainted with that which you do). (Sourate: An Naml_verse: 88)

Dédicace

C'est avec toute mon affection que je dédie ce modeste travail :

*A celui qui m'a donné vie et qui m'as apporté sans cesse son amour,
mon cher père.*

*A ma très chère mère qui m'a apporté sans cesse amour,
soutien et encouragement ; et qu'elle trouve ici l'expression de mes
vives reconnaissances.*

A ma chère femme et ma fille « Yasmine » ma petite princesse.

Mon cher frère Fouzi et ma chère sœur Amani.

*À toute la famille et tous mes amis ! Chacun avec son nom, vous
m'avez beaucoup soutenu ! Et vous allez continuer ... Merci !*

Ouafi

REMERCIEMENTS

Je remercie en premier lieu **ALLAH**, le tout puissant qui a éclairé mon chemin, et ma donné le courage, la force et la patience pour achever ce travail.

Je voudrais remercier vivement mon directeur de thèse, **Pr. Aziez Zeddouri**. Je le remercie de m'avoir proposé ce sujet, de m'avoir encadré et de m'avoir consacré une grande partie de son précieux temps pour la réalisation de ce travail.

Mes remerciements vont aussi à **Pr. Salim Heddami**, mon co-directeur de thèse ; qui n'a jamais cessé à me diriger. Ses conseils et ses orientations, ont permis l'avancement de ce travail. Merci pour la disponibilité et l'intérêt porté à ce travail.

Un grand merci aux membres de jury qui ont accepté d'examiner ce travail : **Dr. Mohammed Salah Belksier** qui m'a fait l'honneur de présider le jury de soutenance. **Dr. Ahmed Ali Zerrouki**, **Pr. Ahmed Hadjaj** et **Pr. Mohammed Chettih** d'avoir accepté d'examiner ce travail.

Je voudrais remercier, à travers cette thèse, **M. Noureddine Meddour** ingénieur Sonatrach. Cette thèse n'aurait jamais avancé sans son aide et son soutien. Merci pour les encouragements.

Je ne peux pas oublier **Dr. Tayeb Kouadria** mon ex-encadreur de l'ingénierat, qu'il trouve ici le témoignage de ma grande reconnaissance et tous les enseignants de université d'Annaba.

Une «montagne» de merci à **Dr. Rabah Kechiched** mon ami et mon frère, j'ai souvent voulu te remercier pour tout ce que tu avais fait pour moi sans jamais avoir trouvé les mots et le moment opportun pour le faire. Tu m'as transmis avec tant de générosité ta passion pour la géologie, et l'art de la recherche scientifique.

Je ne peux oublier mes frères : **Abdellatif Remita « Chakhchoukh »**, **Abdennour Haddane**, **Mohamed-Salah Belksier**, **Nacereddine Slamene**, **Adel Satouh**, **Ammar Amata**, **Abdesslem Brioua**, **Nasro Slamene**, **Mohammed Zighed**, **Hamza Larit**, **Hamza Laouini**, **Bilal Fnazi**, **Abderraouf Remita**, **Boualem Bouselsal**, **Boukhamla Akram**, **Ali Bouamrane**, **Ossama Derdous**, **Zoubir Aissani**, et tous les collègues de départements des sciences de la terre et de l'univers de l'université Kasdi Merbah Ouargla,

Je vous remercie aussi le **Pr. Dobby Abdelmadjid** le doyen de la faculté des hydrocarbures, **Dr. Chetti Djamel-Eddine** le vice doyen et **Pr. Hacini Messaoud** le directeur de laboratoire de géologie de Sahara.

J'adresse mes vifs remerciements mes ex-étudiants de master qui ont contribué efficacement à ce travail, je suis reconnaissant à: **Hamida Touati**, **Nesrine Dahdi**, **Rima Bouhafis**, **Maamri Abdenaceur**, **Abderrezek Hamdat**.

Je remercie toute ma famille, pour le soutien le long de mon parcours, chacun avec son nom qu'ils trouvent ici ma gratitude les plus distinguée. Merci à **Papa** et **Maman**, à **mon frère** et **ma sœur**. Merci à **ma femme** et **ma belle-sœur Nesrine** pour la patience et pour le soutien le long des années de thèse.

الملخص

من أجل التشخيص الجيد لخصائص المكامن البترولية يجب القيام بقياسات تقنية مباشرة مخبريا على عينات حقيقية ، لكن عملية استخراج هذه العينات ليست متاحة دائماً ، بسبب المشاكل التقنية والجيولوجية التي تتم مواجهتها أثناء عملية الحفر. لمعالجة هذا المشكل ، يجب البحث عن أداة فعالة وسريعة لإعادة بناء السحنات الغير مستردة وتقدير بعض الخصائص البتروفيزيائية بدقة مثل المسامية والنفاذية، والتي تتمثل كعناصر أساسية في تقييم الحقل البترولي وقدرته الإنتاجية. تم اقتراح العديد من الصيغ و الدوال الخطية من أجل حل هذا المشكل ، ولكنها تبقى محدودة وغير فعالة كما يمكن أن تسبب في تقديرات خاطئة للاحتياطيات النفط الموجودة فعليا ، إما عن طريق التقليل أو المبالغة في تقديرها ، وبالتالي سيؤثر ذلك على تطوير الحقل ومردوديته. على مدى العقدين الماضيين ، تم اقتراح العديد من الخوارزميات المتعلقة بالذكاء الاصطناعي لحل المشاكل المعقدة في هندسة البترول والغاز.

الهدف من هذه الأطروحة هو إعادة بناء المجالات الغير مستردة وتقدير المسامية والنفاذية الغير متاحة، باستخدام العديد من تقنيات الذكاء الاصطناعي على بيانات التسجيلات الجيوفيزيائية للآبار. تم الأخذ كمثال تطبيقي، المكنم الترياسي الطيني الحصوي السفلي الغير متجانس الواقع في حقل سيف فاطيمة بحوض بركين ،جنوب شرق الجزائر. يمكن تلخيص النتائج الرئيسية لهذه الأطروحة على النحو التالي:

1. بفضل نموذج العصبونات الاصطناعية بطوبولوجيا (4-4-4) تمكنا من إعادة بناء 333 متر من السحنات باستخدام التسجيلات الجيوفيزيائية النووية ، مع ترابط $R^2 = 0.92$.
 2. أعطى تطبيق الخوارزمية الجينية مع الإنحدار الخطي المتعدد نتيجة تقدير المسامية إرتباطاً $R^2 = 0.92$ ، وكذلك نوصي بأخذ الكالبر كمتغير فعال في تقدير المسامية في المكامن الرملية الطينية الغير المتجانسة.
 3. فيما يخص تقدير النفاذية أعطت الشبكة العصبية الاصطناعية فعالية أكثر مقارنة بطرق دعم ناقلات الآلات والغابات العشوائية بترابط $R^2 = 0.92$ ونسبة خطأ $RMSE = 0.0613$ و $MAE = 0.0348$.
- تؤكد نتائج هذه الأطروحة فعالية أدوات الذكاء الاصطناعي في التشخيص الجيد للمكامن البترولية ، بما في ذلك تلك غير المتجانسة ، مما يوفر الوقت و الجهد من ناحية. و تقلل من تكلفة عمليات الحفر والقياسات المخبرية من ناحية أخرى.

الكلمات المفتاحية : السحنات الغير مستردة ، المسامية / النفاذية ، المكنم الغير متجانس ، الذكاء الاصطناعي ،

التسجيلات الجيوفيزيائية ، بركين، الجزائر.

Résumé

Une bonne caractérisation des propriétés du réservoir pétrolier nécessite des mesures directes sur les carottes, cependant ces dernières ne sont pas toujours disponibles à cause des problèmes techniques et géologiques rencontrés durant l'opération du carottage. Pour remédier à cette lacune, le défi se situe à la recherche d'un outil efficace et rapide pour reconstituer les faciès non carottés et pour estimer certains paramètres pétrophysiques avec précision tels que : la porosité et la perméabilité représentant des éléments fondamentaux pour l'évaluation du gisement. Diverses formules ont été suggérées pour la prédiction de ses paramètres, mais elles sont limitées et peuvent entraîner des incertitudes très élevées impliquant de fausses évaluations des réserves en place, soit par une sous-estimation ou surestimation et par conséquent, cela va influencer l'aspect économique lié à l'investissement et le développement du gisement. Au cours des deux dernières décennies, de nombreuses méthodes liées l'intelligence artificielle ont été proposées pour résoudre des problèmes complexes dans l'ingénierie de pétrole.

L'objectif de cette thèse se focaliser sur la reconstitution des faciès non carottés et à l'estimation de la porosité et la perméabilité manquante en utilisant plusieurs techniques de l'intelligence artificielle à l'aide des diagraphies de puits. Le réservoir hétérogène du Trias Argileux Gréseux Inferieur du champ de Sif Fatima bassin de Berkine sud-est de l'Algérie, a été pris comme l'exemple d'application.

Les principaux résultats de cette thèse peuvent se résumer comme ce qui suit :

1. Le modèle MLPNN a permis de reconstruire 333 m de faciès non carottés à partir des diagraphies nucléaires, par une topologie de (4-4-4) avec une performance $R^2=0.92$
2. L'intégration de l'algorithme génétique combinée avec la régression linéaire multiple pour l'estimation de la porosité montre une meilleure corrélation $R^2=0.92$, ainsi on recommande de tenir le Caliper comme un variable très explicatif pour la prédiction de la porosité dans les réservoirs hétérogènes argileux gréseux.
3. La technique des réseaux de neurones est la plus fiable par rapport aux méthodes des machines à vecteurs supports et les forêts Aléatoires utilisées pour estimer la perméabilité avec les performances : $R^2=0.92$; $RMSE= 0.0613$ et $MAE= 0.0348$.

Les résultats de cette thèse confirment l'efficacité des outils de l'intelligence artificielle pour la prédiction des paramètres des réservoirs y compris ceux hétérogènes avec des performances notables ce qui permettent un gain de temps d'une part et elle totalement économique d'autre part, car elle minimise les frais du carottage et les mesures au laboratoire.

Mots-clés: Faciès non carottés ; Porosité / perméabilité ; Réservoir hétérogène. Intelligence artificielle. Diagraphies. Berkine. Algérie.

Abstract

The best characterization of the petroleum reservoirs properties requires direct measurements on the cores; however, the latter are not always available due to various technical and geological problems during coring operation. To solve this issue, the challenge is to seek for effective and fast tool to reconstruct non cored lithofacies as well as estimating with precision some petrophysical parameters such as: porosity and permeability which constitute fundamental elements for the reservoir evaluation. Several formulas from the literature were suggested for predicting the missing parameters, but they are limited and can involve very high uncertainties resulting in false evaluations of the reserves in place if by underestimation or over-estimation that consequently, will have an effect on the economic aspects and will impact investment and the field development. During the two last decades, artificial intelligence methods were largely suggested to solve complex problems in the oil engineering field.

The objective of this thesis belongs in the above-mentioned area and is focused on the reconstitution of the non-cored lithofacies and to estimate missing data of porosity and permeability by the use of several techniques of artificial intelligence based on well logging data. The Lower Clayey Sand Triassic of Sif Fatima oilfield (Berkine, south-eastern of Algeria), which is considered as one of heterogeneous reservoirs, was taken as the example to apply on these techniques.

The main results from this thesis can be summarized as follows:

1. The MLPNN model with (4-4-4) topology allowed to reconstruct 333 m of non-cored lithofacies using only the nuclear logs with a performance of $R^2 = 0.92$.
2. The integration of the genetic algorithm combined with multiple linear regressions for the estimation of the porosity shows a best correlation $R^2 = 0.92$, we also recommend the Caliper as a very illustrative variable for the prediction of the porosity in heterogeneous sandy clayey reservoir.
3. The neural network technique is the most consistent method compared to support vector machines and random forests based-approaches in estimating the permeability displaying the following performance: $R^2=0.92$; RMSE= 0.0613 and MAE= 0.0348.

The Results of this thesis confirm the efficiency of artificial intelligence tools for predicting reservoir parameters even they are heterogeneous and they show high performances. The used techniques allow gain of time on the one hand, and to minimize the cost of coring and laboratory measurements on the other one hand.

Keywords: No-cored lithofacies, Porosity / permeability, Heterogeneous reservoir, Artificial intelligence, Well logging, Berkine, Algeria.

LISTE DES COMMUNICATIONS ET DES ARTICLES

- Ameur-Zaimeche, Ouafi., Zeddouri, Aziez., Heddami, Salim., Kechiched, Rabah., 2020. Lithofacies prediction in non-cored wells from the Sif Fatima oil field (Berkine basin, southern Algeria): A comparative study of multilayer perceptron neural network and cluster analysis-based approaches, *Journal of African Earth Sciences*, Volume 166, 2020, 103826, ISSN 1464-343X, <https://doi.org/10.1016/j.jafrearsci.2020.103826>.
- Ameur-Zaimeche, Ouafi., Kechiched, Rabah., Bouhaf, Rima., Mammeri, Abdelnacer., Hamdat, Abderrazak., Zeddouri, Aziez., 2019. Volume of Clay Estimation Using Artificial Neural Network Case Study: Berkine Basin Southern Algeria. 2nd Conference of the Arabia. *Journal of Geosciences (CAJG'2019)*. 25-28 November 2019, Sousse, Tunisia.
- Ameur-Zaimeche, Ouafi., Zeddouri, Aziez., Kechiched, Rabah., 2019. Multilayer perception neural network-based approach for permeability modelling in TAGI reservoir, case study from Sif Fatima oil field, Algeria. 8th International Symposium on Hydrocarbons and Chemistry SHC8 2019.08-10 April, 2019, Boumerdès, (Algeria).
- Ameur-Zaimeche, Ouafi., Touati, Djemaa., Dahdi, Nersrine., Meddour, Nordine., Kechiched, Rabah., Zeddouri, Aziez., 2019. Development of Artificial Neural Network models for predicting permeability: case study of Sif Fatima oil field, Berkine basin (Southern of Algeria). *International Conference on Artificial Intelligence and Information Technology, ICA2IT'19*, 4-4-6, Mars, 2019, Ouargla, (Algeria).
- Ameur-Zaimeche, Ouafi., Zeddouri, Aziez., Heddami, Salim., Kechiched, Rabah. 2014b. Modélisation et reconstitution des facies non carottés à l'aide des méthodes a base des Réseaux de Neurones Artificiels et Cluster Analysis, cas du réservoir Trias Argileux Gréseux inférieur, champ de Sif Fatima, bassin de Berkine, sud de l'Algérie.. 1^{er} Colloque International sur Hydrocarbures, Energies et Environnement (HCEE'2014), Ouargla, Algeria, Novembre 23-24.
- Ameur-Zaimeche, Ouafi., Zeddouri, Aziez., Kouadria, Taib., Kechiched, Rabah., Belksier, Mohammed Salah. 2014a. Use of Cluster Analysis method in log's data processing: prediction and rebuilding of lithologic facies. *International Conference on Environmental Science and Geoscience (ESG '14)* Venice, Italy, March 15-17, 2014 98 – 101.

LISTE DES FIGURES

Fig.1 - Les différentes zones de cognitives	6
Fig.2 - Représentation d'un neurone biologique	9
Fig.3 - Le neurone formel	9
Fig.4 - Fonction de transfert binaire	10
Fig.5 - Mise en correspondance neurone biologique / neurone artificiel	11
Fig.6 - Modélisation simplifiée d'un neurone artificiel	11
Fig.7 - Perceptron à une seule couche avec fonction seuil	14
Fig.8 - Illustration de regroupement en clusters	16
Fig.9 - Deux types de clustering hiérarchique et non hiérarchique	18
Fig.10 - Discrimination à deux classes, avec une séparatrice linéaire	19
Fig.11 - Exemple d'un arbre de décision	21
Fig.12 - Random Forest (Forêts Aléatoires)	23
Fig.13 - Matrice de confusion	26
Fig.14 - Principe des algorithmes génétiques	27
Fig.15 - Organigramme de l'algorithme génétique	28
Fig.16 - Exemples d'opération de croisement	31
Fig.17 - Exemple de mutation	32
Fig.18 - Schéma du vecteurs de support	34
Fig.19 - Le bon hyperplan optimal	35
Fig.20 - Types de noyau (kernel) de SVM	36
Fig.21 - Situation géographique du bassin de Berkine	40
Fig.22 - La carte géologique de l'Algérie	41
Fig.23 - Les éléments structuraux du bassin de Berkine	44
Fig.24 - Log lithostratigraphique du bassin de Berkine modifié.	46
Fig.25 - Roches sources et migration vers le TAGI	47
Fig.26 - Sources et migrations dans le bassin de Berkine	50
Fig.27 - Carte de situation de la zone d'étude « Sif Fatima »	51
Fig.28 - Plan et directions des dépôts fluviatiles	52
Fig.29 - Colonne stratigraphique du champ de Sif Fatima	53
Fig.30 - Carte Structurale au Toit de TAGI réservoir du secteur Sif Fatima	54
Fig.31 - Coupe sismique du secteur Sif Fatima	55

Fig.32 - Photos de quelques carottes à différents intervalles du réservoir TAGI	57
Fig.33 - Architecture du réseau de neurones élaboré pour reconstituer les faciès	64
Fig.34 - Régression entre les faciès réels et les électrofaciès estimés dans les étapes de l'apprentissage, de validation et le test.	66
Fig.35 - Fonction coût de performance des étapes d'apprentissage (ligne bleue), de validation (ligne verte), et du test (ligne rouge)	67
Fig.36 - Résultat des électrofaciès MLPNN pour le puits P_A	70
Fig.37 - Résultat des électrofaciès MLPNN pour le puits P_B	71
Fig.38 - Résultat des électrofaciès MLPNN pour le puits P_C	72
Fig.39 - Résultat des électrofaciès MLPNN pour le puits P_D	73
Fig.40 - Résultat des électrofaciès MLPNN pour le puits P_E	74
Fig.41 - Histogramme des faciès carottes et électrofaciès MLPNN puits P_A	75
Fig.42 - Histogramme des faciès carottes et électrofaciès MLPNN puits P_B	76
Fig.43 - Histogramme des faciès carottes et électrofaciès MLPNN puits P_C	77
Fig.44 - Histogramme des faciès carottes et les électrofaciès MLPNN, puits A, B, C	78
Fig.45 - Résultat des électrofaciès cluster analysis pour le puits P_A	79
Fig.46 - Résultat des électrofaciès cluster analysis pour le puits P_B	80
Fig.47 - Résultat des électrofaciès cluster analysis pour le puits P_C	81
Fig.48 - Résultat des électrofaciès cluster analysis pour le puits P_D	82
Fig.49 - Résultat des électrofaciès cluster analysis pour le puits P_E	83
Fig.50 - Histogramme des faciès carottes et électrofaciès cluster analysis puits P_A	84
Fig.51 - Histogramme des faciès carottes et électrofaciès cluster analysis puits P_B	85
Fig.52 - Histogramme des faciès carottes et électrofaciès cluster analysis puits P_C	86
Fig.53 - Histogramme des faciès carotte et les électrofaciès cluster analysis dans les puits A,B,C	87
Fig.54 - Comparaison des lithofaciès de carotte avec les résultats (MLPNN et cluster analysis) électrofaciès des puits (A, B, C)	89
Fig.55 - Boîte à moustache des lithofaciès de carotte avec les résultats (MLPNN et cluster analysis) électrofaciès des puits carottés (A, B, C)	90
Fig.56 - Cross-plots de la perméabilité avec les différents paramètres	94
Fig.57 - Cross-plots de la porosité avec les différents paramètres	95
Fig.58 - Cercle de corrélations de l'ACP sur l'ensemble de données	96
Fig.59 - Architecture de réseau de neurone approprié pour estimer la perméabilité	100

Fig.60 - Perméabilité carotte (réelle) par rapport à la perméabilité estimée par RNA	101
Fig.61 - Perméabilité carotte (réelle) par rapport à la perméabilité estimée par SVM	103
Fig.62 - Perméabilité carotte (réelle) par rapport à la perméabilité estimée par Random Forest	104
Fig.63 - Comparaison entre la perméabilité carotte et la perméabilité simulée par les différentes techniques de l'intelligence artificielle	105
Fig.64 - Porosité carotte (réelle) par rapport à la porosité estimée par AG-RLM	108
Fig.65 - Architecture de réseau de neurone approprié pour la prédiction de la porosité	109
Fig.66 - Porosité carotte (réelle) par rapport à la porosité estimée par RNA	111
Fig.67 - Porosité carotte (réelle) par rapport à la porosité estimée par SVM	112
Fig.68 - Porosité carotte (réelle) par rapport à la porosité estimée par Random Forest	113
Fig.69 - Comparaison entre la porosité carotte et la porosité simulée par les différentes techniques de l'intelligence artificielle	114
Fig.70 - Comparaison entre la perméabilité carotte normalisée avec la perméabilité RNA	115
Fig.71 - Comparaison entre la porosité carotte avec la porosité RNA	115

LISTE DES TABLEAUX

Tableau. 1- Différents types de fonctions de transfert pour le neurone artificiel	12
Tableau. 2 - Statistique des paramètres de diagraphies (GR, RHOB, THOR et POTA)	62
Tableau. 3 - Performances des modèles de MLPNN élaborés	66
Tableau. 4 - Les Paramètres retenus des modèles de MLPNN	68
Tableau. 5 - Comparaison des facies carottes et électrofaciès MLPNN du puits P_A	75
Tableau. 6 - Comparaison des facies carottes et électrofaciès MLPNN du puits P_B	75
Tableau. 7 - Comparaison des facies carottes et électrofaciès MLPNN du puits P_C	76
Tableau. 8 - Statistique des facies carottes et électrofaciès MLPNN des puits A, B, C	78
Tableau. 9 - Comparaison des facies carottes et électrofaciès MLPNN des puits A, B, C	78
Tableau. 10 - Comparaison des facies carottes et électrofaciès cluster analysis puits P_A	84
Tableau. 11 - Comparaison des facies carottes et électrofaciès cluster analysis, puits P_B	86
Tableau. 12 - Comparaison des facies carottes et électrofaciès cluster analysis, puits P_C	85
Tableau. 13 - Statistique des facies carottes et électrofaciès cluster analysis des puits A, B, C	87
Tableau. 14 - Comparaison des facies carottes et électrofaciès cluster analysis dans les puits (A,B,C)	87
Tableau. 15 - Comparaison des facies carottes et (MLPNN ; cluster analysis) électrofaciès P-A	88
Tableau. 16 - Comparaison des facies carottes et (MLPNN ; cluster analysis) électrofaciès P-B	88
Tableau. 17 - Comparaison des facies carottes et (MLPNN ; cluster analysis) électrofaciès P-C	88
Tableau. 18 - Statistiques de porosité, perméabilité et des données de diagraphies	93
Tableau. 19 - Matrice de corrélation (porosité et la perméabilité carotte) et les enregistrements diagraphiques	94
Tableau. 20 - Performances des modèles de réseaux de neurones élaborés	101
Tableau. 21 - Critères de performance de la perméabilité pour chaque méthode	105
Tableau. 22 - Paramètres du modèle de (AG-RLM)	107
Tableau. 23 - Performances et architectures des modèles testés	110
Tableau. 24 - Critères de performance de la porosité pour chaque méthode	114

SOMMAIRE

DEDICACE	I
REMERCIEMENT	II
المُلخَص	III
RESUME	IV
ABSTRACT	V
LISTE DES COMMUNICATIONS ET DES ARTICLES	IV
LISTE DES FIGURES	IX
LISTE DES TABLEAUX	X
INTRODUCTION GENERALE	1

CHAPITRE I : SYNTHÈSE BIBLIOGRAPHIQUE SUR LES METHODES DE L'INTELLIGENCE ARTIFICIELLE

1. Généralités sur l'Intelligence Artificielle	4
1.1. L'intelligence.....	4
1.2. La cognition.....	5
1.3. Intelligence artificielle.....	6
1.4. Historique de l'intelligence artificielle.....	7
1.5. Application en sciences de de la terre.....	7
2. Réseaux de Neurones Artificiels	8
2.1. Les réseaux de neurones biologiques.....	8
2.2. Le neurone artificiel.....	10
2.3. Processus d'apprentissage des réseaux de neurones	12
2.4. Perceptron simple et perceptron multicouche	13
3. Cluster Analysis	15
3.1. Définition.....	16
3.2. Principe général.....	17
3.3. Les exigences de clustering.....	17
3.4. Les types de clustering.....	17
3.5. Les algorithmes de clustering.....	19
4. Random Forest (Forêts Aléatoires)	21
4.1. Arbre de décision.....	21
4.2. Random forest (Forêts Aléatoires).....	22
4.3. La dimension de modèle.....	26
5. Algorithmes Génétiques	26
5.1. Concepts et principe des algorithmes génétiques.....	26
5.2. La mise en œuvre d'un algorithme génétique.....	28
5.3. Vocabulaires des algorithmes génétiques.....	28
5.4. Le codage.....	29
5.5. La population initiale.....	29
5.6. Fonction d'évaluation (fitness).....	30
5.7. Les opérateurs de base d'un AG.....	30

5.8. La convergence d'un AG.....	32
6. Support Vector Machines (SVM).....	33
6.1. Mise en œuvre des SVM.....	34
6.2. Support Vector Machines.....	34
6.3. Qu'est-ce qu'un hyperplan?.....	34
6.4. Comment trouver le bon hyperplan?.....	35
6.5. Avantages et les inconvénients.....	35
6.7. Utilisations SVM.....	36
6.8. Type de noyau.....	36
6.9. Applications de SVM dans la vie.....	36

CHAPITRE II : GEOLOGIE REGIONALE ET LOCALE DE LA ZONE D'ETUDE

1. Géologie Régionale du Bassin de Berkine.....	39
1.1. Situation géographique.....	39
1.2. Cadre géologique régionale.....	40
1.3. Evolution tectonique.....	41
1.4. Géodynamique du bassin.....	43
1.5. Stratigraphie.....	44
1.6. Le système pétrolier du bassin de Berkine.....	47
2. Géologie Locale de Champ de Sif Fatima.....	50
2.1. Situation géographique.....	50
2.2. Synthèse géologique et sédimentologique.....	50
2.3. Tectonique locale.....	54
2.4. Le potentiel pétrolier de la zone d'étude.....	55
3. Caractéristique Pétrophysique de Réservoir TAGI de la Zone Sif Fatima....	56
3.1. Analyse des carottes.....	56
3.2. Synthèse pétrophysique du réservoir TAGI.....	57
3.3. Historique des travaux réalisés sur le bassin de Berkine.....	58

CHAPITRE III : PREDICTION DES FACIES NON CAROTTES

1. Préparation de la Base des Données.....	62
2. Elaboration du Modèle des Réseaux de Neurones Artificiels.....	63
2.1. L'architecture du réseau de neurones.....	64
2.2. Amélioration des performances du réseau.....	64
2.3. Critères de validation.....	65
3. Application de la Méthode de Classification Cluster Analysis.....	68

4. Prédiction des Electrofaciès par les Réseaux de Neurones (MLPNN)	69
4.1. Comparaison qualitative entre faciès carotte et électrofaciès MLPNN...	69
4.2. Comparaison quantitative entre faciès carotte et électrofaciès MLPNN.....	74
4.3. Synthèse des résultats de comparaison les faciès carotte et les électrofaciès MLPNN de l'ensemble des puits A, B et C	77
5. Prédiction des Electrofaciès par la Méthode de Cluster Analysis	79
5.1. Comparaison qualitative entre faciès carotte et électrofaciès cluster analysis.....	79
5.2. Comparaison quantitative entre faciès carotte et électrofaciès cluster analysis.....	84
5.3. Synthèse des résultats de comparaison faciès carotte et électrofaciès cluster analysis de l'ensemble des puits A, B et C.....	86
6. Comparaison Entre les Résultats des Electrofaciès MLPNN et de Cluster Analysis	88

CHAPITRE IV: PREDICTION DE LA PERMEABILITE ET LA POROSITE

1. Méthodologie	92
1. 1. Analyses et préparation de la base de données	92
2. Prédiction de la Perméabilité par des Différentes Techniques de l'Intelligence Artificielle	99
2.1. Application du réseau de neurones artificiels	100
2.2. Application de Support Vector Machine (SVM)	102
2.3. Application de la méthode des Forêts Aléatoires (Rondom Forest).....	103
2.4. Comparaison entre la perméabilité carotte et la perméabilité simulée par les différentes techniques de l'intelligence artificielle.....	104
3. Prédiction de la Porosité par des Différentes Techniques de l'Intelligence Artificielle	106
3.1. Application du modèle Hybride d'Algorithmes Génétiques et Régression Linéaire Multiple (AG-RLM).....	107
3.2. Application des réseaux de neurones artificiels.....	109
3.3. Application de Support Vector Machine (SVM).....	112
3.4. Application de la méthode des Forêts Aléatoires (Rondom Forest).....	113
3.5. Comparaison entre la porosité carotte et la porosité simulée par les différentes techniques de l'intelligence artificielle	114
4. Généralisation des Modèles de Porosité et de Perméabilité	115
CONCLUSION GENERALE	118
BIBLIOGRAPHIE	121
ANNEXES	129

INTRODUCTION GENERALE

INTRODUCTION GÉNÉRALE

La conjoncture actuelle du marché pétrolier qui se manifeste par une chute des cours de brute, d'une instabilité des cours face à une influence liée à beaucoup de facteurs (politiques, géostratégiques, sanitaires « coronavirus »..) d'une part et la forte production mondiale qui cesse pas à s'accroître depuis le « boom » pétrolier des réservoirs non-conventionnels surtout ceux du schiste américain et le gonflement des stocks de brut. Face à cette situation, l'économie de l'Algérie qui reste dépendante exclusive aux recettes des hydrocarbures de plus de 90 %, a été fortement touchée par cette conjoncture, et comme d'autres pays producteurs au sein de l'OPEP doivent agir efficacement pour limiter le coût de production afin de garantir une marge bénéficiaire respectable et ceci avec, entre autres, le bon contrôle des coûts de production en particulier dans les phases d'exploration et de développement, dont lesquelles beaucoup d'opérations coûteuses de forage et de carottage ainsi que la caractérisation des paramètres interviennent, alors il est nécessaire de travailler afin de limiter les dépenses. Pour cela, l'emploi des nouvelles technologies innovantes pourrait limiter la facture consacrée d'une part et intervenir efficacement pour résoudre les problèmes liés à cette activité tels que : les carottes manquantes et la mauvaise récupération, les paramètres non mesurés et autres.

Il est clair que le meilleur moyen pour une bonne caractérisation et évaluation des réservoirs pétroliers sont les mesures directes sur les carottes ; cependant la récupération de ses derniers n'est pas toujours intégrale, à cause des problèmes techniques et géologiques rencontrés surtout lors de l'opération du carottage et sa limitation en considérant le nombre des puits carottés dans le champ (Dubois et al, 2007). D'autant plus, pendant la modélisation du réservoir, l'identification des lithofacies s'avère une phase très importante dans l'élaboration des modèles statiques et dynamiques. D'autre part pour qu'un réservoir soit économiquement rentable, deux propriétés essentielles doivent bien caractérisées et évaluées avec précision. Il s'agit de la capacité de stockage des hydrocarbures et de la transmissibilité des fluides (porosité-perméabilité). Ses paramètres sont jusqu'à ce jour, mesurés directement sur les carottes qui constituent le meilleur moyen pour avoir des données fiables.

Une autre alternative représentée par prédiction où diverses techniques et formules empiriques ont été suggérées dans la littérature pour la prédiction de ses propriétés à savoir par exemple : Wyllie–Rose, Timur, Kozeny-Carman à partir des

diagraphies de puits (Jamialahmadi & Javadpour, 2000; Lim, 2005; Mohaghegh, 2005), mais elles restent limitées et peuvent entraîner des incertitudes très élevées impliquant des fausses évaluations des réserves en place, soit par une sous-estimation ou surestimation et par conséquent, cela va influencer sur l'aspect économique lié à l'investissement et le développement du gisement.

Au cours des deux dernières décennies, les techniques de l'intelligence artificielle telles que l'informatique neuronale, les arbres de décision et les génétiques algorithmes sont devenus des outils de plus en plus puissantes, et ils font une révolution dans l'ingénierie de pétrole et de gaz (Asoodeh & Bagheripour, 2012; Zargar et al., 2014 ; Dramsch, 2020), par ses capacités de résoudre les problèmes complexes survenant pendant les mesures, par exemple, une des logs incomplets, un stockage de données inapproprié ou des erreurs de mesure, des données manquantes dans les diagraphies ..etc.

Objectifs de la thèse

L'objectif de la thèse se focalise essentiellement sur la reconstitution des facies non carottés et à l'estimation de la porosité et la perméabilité manquante, en utilisant plusieurs techniques de l'intelligence artificielle (*les réseaux de neurones artificiels, support vector machine, les algorithmes génétiques, les forêts aléatoires, et cluster analysis*). Le principe repose sur une transformation des données aux informations et les informations en connaissances pour une bonne évaluation de notre réservoir.

Ces techniques vont employer les diagraphies de puits pour pallier le manque d'information, dans le réservoir du trias argileux gréseux inférieur (TAGI), champ de Sif Fatima nord-est du bassin de Berkine. Ce dernier est un réservoir de type fluviatile très hétérogène et renferme des quantités importantes des hydrocarbures.

Les roches réservoirs sont constituées par les séries basales du mésozoïque, elles sont marquées par les dépôts de nature fluviatile qui se développent le long du bassin. Il a deux principales séquences : le TAGI inférieur et le TAGI supérieur et moyen où les formations des chenaux et les méandres prédominent.

Organisation de la thèse

Cette thèse est composée de quatre chapitres qui sont organisés de la façon suivante:

Chapitre I : ce chapitre présente les différentes techniques de l'intelligence artificielle qui sont utilisés pour résoudre les problèmes posés à savoir : les facies non-carottés et l'estimation de la porosité et la perméabilité manquante.

Chapitre II : ce chapitre consiste la présentation de la région d'étude ; le cadre géologique régional du bassin de Berkine, la géologie locale du champ de Sif Fatima et la caractérisation sédimentologique du réservoir Trias Argileux Gréseux Inférieur TAGI.

Chapitre III : la présentation des résultats de la reconstitution des facies non carottés par les réseaux de neurones artificiels et comparés par la méthode de classification cluster analysis, à partir les diagraphies nucléaires et la description des carottes.

Chapitre IV : l'estimation de la porosité et la perméabilité par les différentes techniques de l'intelligence artificielle à l'aide des diagraphies de puits disponibles, ainsi que les données réels de perméabilité et de porosité.

Ce travail s'achève par une conclusion générale présentant les principaux résultats obtenus et une discussion portant sur les perspectives d'étude.

Chapitre I

Synthèse Bibliographique sur les Méthodes de l'Intelligence Artificielle

Chapitre I. Synthèse bibliographique sur les méthodes de l'intelligence artificielle

Introduction

Ces dernières années, l'intelligence artificielle est devenu un outil interdisciplinaire de plus en plus important qui a fait progresser plusieurs domaines scientifiques, tels que la chimie (Schütt et al., 2017), la biologie (Ching et al., 2018), la médecine (Shen et al., 2017) et la pharmacie (Kadurin et al., 2017). D'après la bibliographie, l'intelligence artificielle a été appliquée pratiquement dans tous les aspects de géoscience. La plupart des sous-disciplines de la géoscience ont fait l'objet d'un examen de ces techniques pour résoudre les problèmes complexes; la télédétection a été l'un des premiers à adopter (Lary et al., 2016), , l'hydrogéophysique (Shen, 2018), sismologie (Kong et al., 2019), interprétation sismique (Wang et al., 2018) et géochimie (Zuou et al., 2019). Le changement climatique, en particulier, a fait l'objet d'un examen approfondi de l'impact potentiel de méthodes de l'intelligence artificielle pour la modélisation. De nombreuses méthodes et concepts mathématiques et statistiques, y compris la règle de Bayes (Bayes, 1763), les moindres carrés (Legendre, 1805) et les modèles de Markov (Markov, 1906, 1971), sont appliqués à l'apprentissage automatique.

Dans le domaine de l'ingénierie de pétrole a connu une large utilisation des techniques de l'intelligence artificielle pour résoudre des problèmes complexes dans la quasi-totalité de la chaîne d'exploitation de l'industrie pétrolière ainsi qu'aux différentes applications sur les données sismiques (Dramschi, 2020-2019 ; Diersen, et al., 2011), la géochimie (Sun et al., 2009), le forage (Kannan et al., 2014), le well testing (Adibifard et al., 2014) et les diagraphies (Salehi et al., 2017). Dans ce chapitre on va présenter une synthèse bibliographique des différentes technique de (IA) utilisés pour résoudre les problèmes de non récupération des carottés et les paramètres pétrophysique manquante dans le champ le champ de Sif Fatima, bassin de Berkine sud-est de l'Algérie.

1. Généralités sur l'Intelligence Artificielle

1.1. L'intelligence

Malgré une longue histoire de recherche et de débats, il n'y a pas une définition standard de l'intelligence. Cela a conduit certains à croire que l'intelligence peut être décrite de manière approximative, mais ne peut pas être entièrement définie (Legg &

Hutter 2007a), la définition la plus universelle selon (Legg & Hutter 2007b), on peut définir l'intelligence étymologiquement le mot « intelligence » est dérivé du mot latin « *intelligentia* » qui est signifié la faculté de comprendre et plus exactement, c'est la aptitude à lier les éléments entre eux, elle est l'ensemble des capacités mental qui permettant de comprendre l'entourage de l'homme, et de découvrir des relations nous permettant d'aboutir à une connaissance conceptuelles et rationnelles celle-ci étant en opposition avec les sensations et émotion (Binet, 1911). La compréhension est le résultat d'un système codage diversifié qui, par maîtrise du langage permet un raisonnement complexe qui nous aide à faire des relations entre éléments (Bingham, 1937). C'est également la définition de la faculté d'adaptation, nous laissant de gérer de nouvelles situations et aussi gérer la capacité à analyser les informations pour aboutir un objectif. Afin d'atteindre ces l'objectifs, l'intelligence fera intervenir de plusieurs facultés cognitives qui sont basées sur notre connaissance du cerveau humain, et inspirées de ces processus cognitive.

1.2. La cognition

Un processus cognitif, est un cheminement par lequel un système analyse une information en y répandant par une action. Ce traitement est le processus qui permet analyser l'information et l'intégrée dans la base de connaissance du système, et ceci dépendant du mode de traitement et du niveau d'élaboration (Binet, 1911), elle rassemble l'ensemble des processus mentaux, c'est-à-dire que tout est dissociable de la perception et de l'action (Fig. 1). Si nous nous en tenons à la définition anglaise de la cognition c'est, « *Get a knowledge of* », cela englobe toutes les capacités qui sont liées au savoir : l'acquisition et mémorisation des informations, la rétention du mémoire et la manipulation conscience, l'apprentissage, jugement, raisonnement, et la langage, etc. alors il existe une grande différence fondamentale entre la cognition et l'intelligence. L'intelligence est une propriété, que l'on observe chez l'être humain mais également sur d'autres animaux ou insectes in (Panzoli, 2008).

Les processus de cognitifs ont été regroupés comme suit:

- Perception, sensation, attention.
- Catégorisation, reconnaissance, raisonnement et prise de décision.
- Représentation, langue et mémoire.

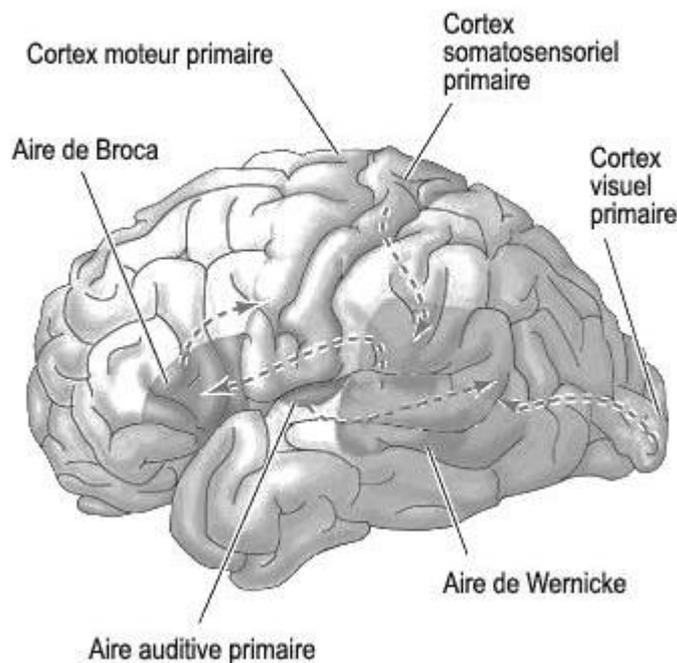


Fig. 1 - Les différentes zones de cognitives.

(https://lecerveau.mcgill.ca/flash/a/a_12/a_12_cr/a_12_cr_con/a_12_cr_con.html)

1.3. Intelligence artificielle

C'est une discipline scientifique relative au traitement des connaissances et au raisonnement dans le but de permettre à une machine d'exécuter des fonctions normalement associées à l'être humain (Albus, 1991) L'intelligence artificielle tente de reproduire les processus cognitifs humains dans le but de réaliser des actions « intelligente ». Ceci comme la création des programmes informatiques qui vont donner des tâches qui sont jusqu'au présent accomplies de manière plus satisfaisante par un être humain, parce que ils ont les processus mentaux de haut niveau tels que :

- L'apprentissage perceptuel.
- Le raisonnement critique et l'organisation de la mémoire.

La norme de l'ISO numéro2382-28 a défini l'intelligence artificielle entant que la « capacité d'une unité fonctionnelle à exécuter des fonctions généralement associées à l'intelligence humaine, telles que l'apprentissage et le raisonnement». Qui qualifiée aux prochaines révolutions dans le domaine de l'informatique, l'intelligence artificielle est au centre de tous les sujets d'actualité, il s'avère obligatoire de bien définir cette technologie.

1.4. Historique de l'intelligence artificielle

L'intelligence artificielle est un domaine qui est bien apparu dernièrement, mais elle a une soixantaine d'années environ. Ce domaine est très complexe qui mélange les sciences, les théories et les techniques a pour ambition d'imiter les capacités intelligentes et cognitives d'un être humain. Elle a été commencée en 1940-1960, l'idée de faire le lien entre le mode de fonctionnement des machines et des êtres organiques. L'objectif de Norbert Wiener, qui est le pionnier dans la cybernétique, était d'allier différents domaines qui sont les mathématiques, l'électronique et l'automatisation tout en conjuguant la dimension animale. Ce projet avait déjà été entamé par les scientifiques Warren McCulloch & Walter Pitts 1943 ont créé le tout premier neurone biologique modélisé mathématiquement par un neurone formel.

1.5. Application en sciences de de la terre

Entre 1950 et 2020, beaucoup de choses ont changé. Les ressources informatiques sont désormais largement disponibles en tant que matériel et logiciel, le calcul de haute performance étant accessible. Haute qualité des logiciels pour l'apprentissage automatique sont largement disponibles via le mouvement des logiciels libres et open-source, en concurrence pour l'utilisation de leurs frameworks d'apprentissage automatique open source (Tensorflow, Pytorch, CNTK2) et des développements indépendants atteignant de larges applications telles que scikit-learn (Pedregosa et al., 2011) et xgboost (Chen & Guestrin, 2016).

Les modèles d'apprentissage automatique, statistiques et mathématiques ont une longue histoire en géoscience. Des modèles de Markov ont été utilisés pour décrire la sédimentologie dès les années 1970 (Schwarzacher, 1972) et l'utilisation des k-moyennes en géoscience dès 1964 (Preston et al., 1964). En géophysique, les applications des NN pour effectuer une déconvolution sismique étaient publiés dans les années 80 Zhao et al. (1988). Les premières méthodes basées sur les arbres étaient principalement utilisées géologie économique et géophysique d'exploration pour la cartographie de prospectiviste avec Les arbres de décisions (DT) (Newendorp, 1976; Reddy et al., 1991). SVM a été appliqué dès le début à classification AVO Li et al., 2004 et délimitation des faciès géologiques pour l'analyse hydrologique (Tartakovsky, 2004).

L'intelligence artificielle a été examinée dans un contexte géophysique. De bonnes publications de (IA) en géoscience appliquent les NN à des problèmes géophysiques. Le traitement sismique en particulier se prête à l'exploration des NN en

tant qu'approximateur fonctionnel général (Hornik). Les réseaux neuronaux ont les premières applications des principales dans le traitement des données sismiques et l'analyse. (Drams, 2020; Zhao & coll. 1988) utilisent un NN pour effectuer une déconvolution sismique.

2. Les réseaux de Neurones Artificiels

Les réseaux de neurones artificiels sont des méthodes intelligente artificielle, leurs évolutions a passé par la manière dont les humains tentent constamment d'imiter la nature et de reproduire des modes de raisonnement et de similarités. Depuis cinquante ans, les informaticiens se sont inspirés de la biologie pour produire des structures capables de résoudre divers problèmes complexes.

La nature a forgé depuis longtemps le premier processeur (l'A.D.N1) et les mécanismes d'optimisation puissants, ainsi des structures de calcul massivement parallèles possédant de bonnes capacités d'adaptation comme notre cerveau. Alors, la centaine de milliards de neurones qui sont interconnectés chacun a parfois plus de dix mille autres de manière tridimensionnelle de notre cerveau, ce qui a attiré l'attention des informaticiens (Yuste & Tank ,1996 in Romain ,2003).

2.1. Les Réseaux de neurones biologiques

Les réseaux de neurones artificiels sont inspirés du processus biologiques pour le traitement de l'information, en particulier le système nerveux et de son unité de base, le neurone. Les signaux sont reproduits comme la différence de potentiel entre l'intérieur et l'extérieur des cellules. Les composants de cellules neuronales sont représentés par la figure (2). Dendrites introduisent des signaux en provenance d'autres neurones dans le corps cellulaire ou soma, éventuellement en multipliant chaque signal d'entrée par un coefficient de pondération. Dans le soma, la capacité de la cellule intègre les signaux qui s'accumulent dans l'axone monticule (axon hillock). Une fois que le signal est dépassé un certain seuil de la cellule, un signal de potentiel d'action est alors transmis à travers l'axone (in Tarek ,2004).

L'axone se connecte via les synapses avec les dendrites des neurones suivants. Les synapses fonctionnent grâce à la décharge de neurotransmetteurs dans les lacunes intercellulaires, et peuvent être soit excitateurs (tendant à tirer le neurone) ou inhibiteurs (tendant à empêcher l'activation du prochain neurone).

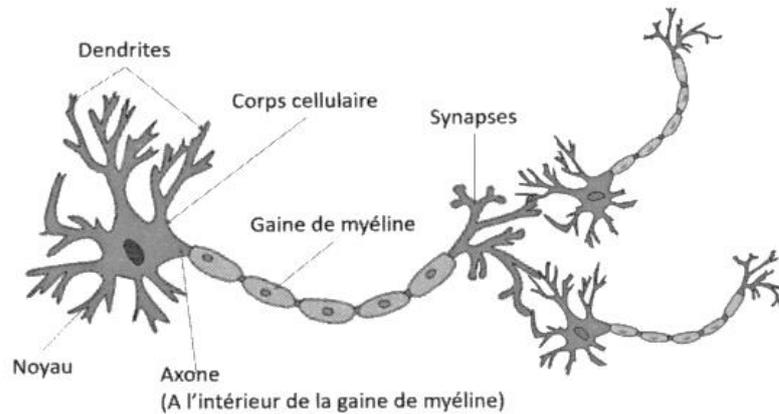


Fig. 2 - Représentation d'un neurone biologique
<https://commons.wikimedia.org/wiki/File:Neuron.jpg>

a) Le Modèle neurone formel

Un neurone formel est une fonction algébrique non linéaire et bornée, avec la valeur à la sortie dépend des paramètres qui sont appelés les coefficients ou poids. Les variables de cette fonction sont généralement appelées "entrées" du neurone, et la valeur de la fonction est appelée la "sortie". Avant tout un neurone est un opérateur mathématique par lequel on peut calculer la valeur numérique avec quelques lignes de logiciel. On a pris l'habitude de symboliser graphiquement un neurone comme présenté sur la figure.3 (McCulloch et Pitts , 1943).

Si aucune synapse inhibitrice n'est active, la fonction du neurone est simple: le neurone ajoute son entrée synaptique et vérifie si la somme atteint ou dépasse son seuil. On dit alors que le neurone devient actif, sinon le neurone est inactif (Fig. 3). Voyons tout cela en détail sur un exemple. Le neurone reçoit les entrées $x_1, x_2, x_3 \dots x_n$.

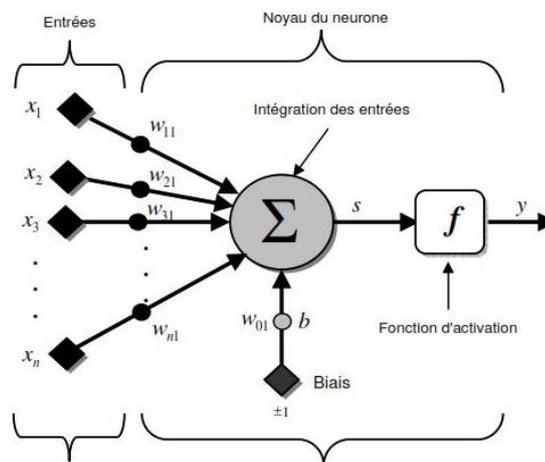


Fig. 3 - Le neurone formel

<http://informatique.coursgratuits.net/methodes-numeriques/reseaux-de-neurones-formels.php>

X_1, X_2, X_3, X_n sont les entrées, ou encore les excitations.

W_1, W_2, W_3 sont appelés les coefficients synaptiques ou poids synaptiques, actuellement nous utilisons des poids différents pour $W_1, W_2, W_3 \dots$. Les poids peuvent même être "corrigés" dans le cas de l'apprentissage. le potentiel p est donné par

$$p = x_1 * W_1 + x_2 * W_2 + x_3 * W_3 \dots + x_n * W_n$$

La fonction de sortie est $s = f(p)$ est la fonction de transfert. Le neurone créé par [McCulloch et Pitts](#) était un neurone utilisant une fonction binaire (Fig. 4).

C'est-à-dire si $p > a$ alors $S = 1$ sinon $S = -1$.

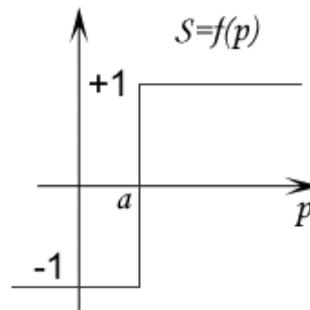


Fig. 4 - Fonction de transfert binaire ([Dreyfus, 1998](#)).

D'autres fonctions sont utilisées en particulier les fonctions sigmoïdes. Alors on peut ensuite créer un réseau suivant les besoins.

a) Les Modèles mathématiques

Les réseaux de neurones biologiques peuvent se réaliser facilement un certain nombre d'applications telles que le traitement du signal, la reconnaissance de formes, l'apprentissage par l'exemple, la mémorisation, la généralisation. Ces applications sont, malgré tous des efforts déployés en algorithmique et en intelligence artificielle, à la limite des possibilités actuelles ([inTouzet, 1992](#)).

2.2. Le Neurone artificiel

a) Structure

La figure ci-dessous (Fig. 5) montre la structure d'un neurone artificiel, pour chaque neurone artificiel est un processeur élémentaire, il reçoit un nombre des variables d'entrées qui sont provient des neurones amonts, à chacune de ces entrées sont associés un poids w abréviation de weight représentatif de la force de la connexion. Chaque processeur élémentaire a une sortie unique, qui se ramifie par la suite pour alimenter un nombre variable des neurones en aval. Dans chaque connexion est associé un poids ([Touzet, 1992](#)).

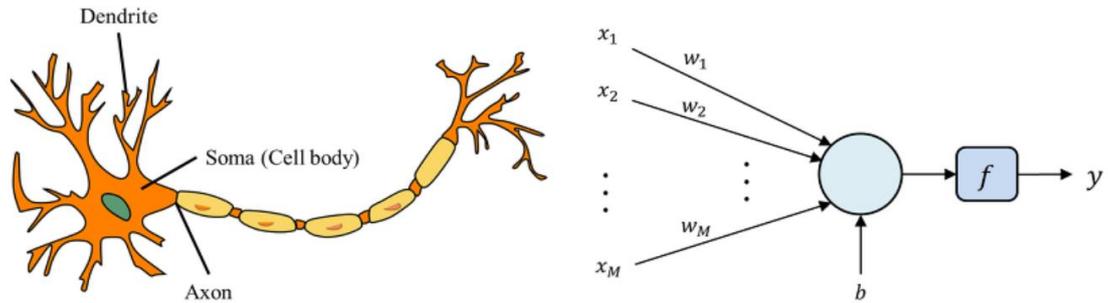


Fig. 5 - Mise en correspondance neurone biologique / neurone artificiel (<https://smartstuartkim.wordpress.com/2019/01/27/history-of-neural-networks-1-perceptron/>)

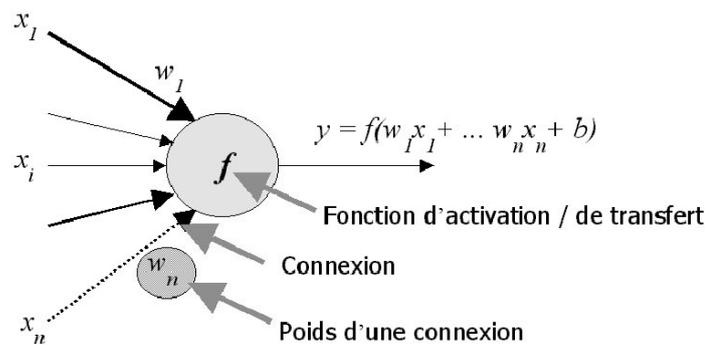


Fig. 6 - Modélisation simplifiée d'un neurone artificiel (Touzet, 1992).

Pour le neurone d'indice i , les entrées sur celui-ci sont de poids w_{ij} alors que les connexions aval sont de poids w_{ji} .

b) Comportement et fonction de transfert

On distingue deux phases. La première est habituellement le calcul de la somme pondérée des entrées (n) selon l'expression suivante :

$$n = \sum (w_i \cdot e_i) \quad (1)$$

Sur la base de cette valeur, la fonction de transfert calcule la valeur de l'état neuronal. Cette valeur sera transmise aux neurones en aval. Il existe de nombreuses formes possibles de la fonction de transfert selon Mar (2004). Les plus courants sont indiqués dans le (tableau 1). Contrairement, il est noté aux neurones biologiques dont l'état est binaire, la plupart des fonctions de transfert sont continues, fournissant l'intervalle $[0, +1]$ (ou $[-1, +1]$).

Tableau. 1 : Les types de fonctions de transfert pour le neurone artificiel (Marc, 2004).

Nom de la fonction	Relation d'entrée –sortie	Forme
1. Seuil	$a = 0$ si $n < 0$ $a = 1$ si $n \geq 0$	
2. Seuil symétrique	$a = -1$ si $n < 0$ $a = +1$ si $n \geq 0$	
3. Linéaire	$a = n$	
4. Linéaire saturé	$a = 0$ si $n < 0$ $a = n$ si $0 \leq n \leq 1$ $a = 1$ si $n > 1$	
5. Linéaire saturé symétrique	$a = -1$ si $n < -1$ $a = n$ si $-1 \leq n \leq 1$ $a = 1$ si $n > 1$	
6. Linéaire positive	$a = 0$ si $n < 0$ $a = n$ si $n \geq 0$	
7. Sigmoide	$a = 1/[1 + \exp^{-n}]$	
8. Tangente hyperbolique	$a = [e^n - e^{-n}] / [e^n + e^{-n}]$	
9. Compétitive	$a = 0$ si n maximum $a = 1$ si autrement	

2.3. Processus d'apprentissage des réseaux de neurones

2.3.1. Définition

Les propriétés désirables et la plus fondamentale pour les réseaux de neurones, c'est la capacité d'apprendre de son environnement, d'améliorer sa performance à travers un processus d'apprentissage. Alors *qu'est-ce donc que l'apprentissage ?* Malheureusement, il n'existe pas de définition générale, universellement bien adoptée, car ce concept touche à trop de notions distinctes qui dépendent du point de vue que l'on adopte. L'apprentissage : est un processus itératif et dynamique qui permet de modifier les paramètres d'un réseau en réaction avec les stimuli, qu'il reçoit de son environnement. Le type d'apprentissage dépend de la manière dont les paramètres changent (Hinton, 1989).

2.3.2. Type d'apprentissage

Selon l'intervention ou non d'un superviseur, on distingue trois types d'apprentissage :

- **L'apprentissage supervisé :** Ce mode est le plus courant, l'utilisateur dispose d'un comportement de références précis qu'il désire faire apprendre au réseau.

Schématiquement, il consiste à évaluer les poids synaptiques minimisant l'erreur sur une base d'apprentissage, cette base comprend un ensemble d'observations pour lesquelles on connaît à la fois les entrées et les sorties souhaitées.

- **L'apprentissage semi- supervisé** : Cet apprentissage est identique au précédent dans la mesure où il se base sur la présence d'un concepteur, cependant la valeur exacte de la sortie n'est pas disponible, en général la seule information disponible est un signal d'échec ou de succès.
- **L'apprentissage non supervisé** : Ce mode d'apprentissage est moins intuitif, il correspond au cas où l'on ne dispose pas de bases d'apprentissage, par exemple lorsqu'on ne sait pas a priori déterminer ponctuellement si une sortie est ou non valable. L'apprentissage repose alors sur un « critère interne » de conformité du comportement du réseau par rapport à des spécifications générales et non sur des observations (Hinton, 1989) simple et perception multicouche.

2.4. Perceptron simple et perception multicouche

2.4.1. Le Perceptron Simple

Mis au point à partir du milieu des années 50 (57-61) par Rosenblatt (Rosenblatt, 1962). Le perceptron veut un modèle de l'activité perceptive qu'il s'agit une association des configurations présentées par les entrées (fig. 7). Présenté par une seule couche de S neurones totalement connectée sur un vecteur P -de- R -entrée. L'objectif est d'entraîner le réseau jusqu'à ce que la valeur obtenue en sortie soit égale à la valeur désirée (si on est dans le cadre d'un apprentissage supervisé) (Corsini, 2005). Pour cette raison on utilise la règle de Hebb qui est basée sur le renforcement des connexions entre deux cellules si elles répondent et de diminuer similairement cette valeur sinon. On pose:

$$w_{ij} = w_{ij} + \eta (d_i - o_i) s_j \quad (2)$$

en plus, si on met le temps plus explicite dans l'équation

$$w_{ij}(t+1) = w_{ij}(t) + \eta(t) (d_i - o_i(t)) s_j(t) \quad (3)$$

Dans laquelle $o_i(t)$ qui désigne la valeur calculée (au temps t) par le réseau sur la sortie i , quand que d_i correspond à la valeur désirée ; $s_j(t)$ correspond au signal transitant sur la connexion de j à i de poids $w_{ij}(t)$; $\eta(t)$ étant le pas (ou coefficient) d'apprentissage au temps t . Ce pas sert à moduler l'impact de l'erreur sur la connexion, η le taux d'apprentissage.

Pour la sortie du réseau, elle peut prendre uniquement deux valeurs différentes selon le niveau d'activation du neurone : -1 lorsque ce dernier est strictement inférieur à 0 ; +1 dans le cas contraire.

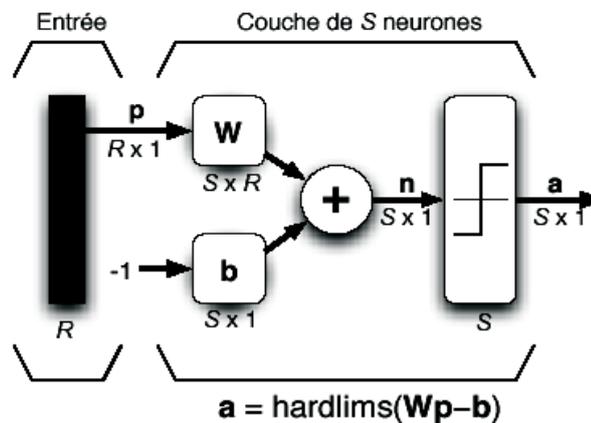


Fig. 7 - Perceptron à une seule couche avec fonction seuil

(<http://informatique.coursgratuits.net/methodes-numeriques/reseaux-de-neurones-formels.php>)

2.4.2. Algorithme de traitement

Selon Dreyfus, 1998 ; l'algorithme de traitement consiste à :

✓ Initialisation

On mène les poids initiaux w_1, w_2, \dots, w_n ainsi que le seuil θ à des valeurs aléatoires de l'intervalle $[-0,5, 0,5]$. Et on met le taux d'apprentissage à une petite valeur positive.

✓ Activation

L'activation du perceptron en appliquant les intrants $x_1(p), x_2(p), \dots, x_n(p)$ et les extrants désirés $Y_d(p)$. Calcule des extrants actuels à l'itération $p = 1$.

$$Y(p) = \text{étage} \left[\sum_{i=1}^n x_i(p) w_i(p) - \theta \right] \quad (4)$$

où n c'est le nombre des signaux intrants, et étage est la fonction d'activation par étage.

✓ Entraînement des poids

Mettre à jour les poids du perceptron

$$w_i(p+1) = w_i + \Delta w_i \quad (5)$$

où $\Delta w_i(p)$ est la correction de poids à l'itération p . Pour la correction des poids, elle est

calculée par la loi delta :

$$\Delta w_i = \alpha \times x_i(p) \times e(p) \quad (6)$$

où α est le taux d'apprentissage et $e(p)$ l'erreur à l'itération p (la différence entre l'extrait désiré et l'extrait actuel du perceptron)

✓ **Itération**

Augmentation de p de 1, et le retour à l'étape 2 et répéter le procédé jusqu'à la convergence.

❖ **Adaline et Widrow-Hoff**

Dans la section précédente, on a traité le cas du perceptron simple, dont les neurones utilisent une fonction de transfert de type s'appelé 'seuil'.

2.4.3. Perceptron multicouche et rétro propagation de l'erreur

❖ **Définition**

Les perceptrons Multicouches sont des réseaux dont les neurones sont organisés en couches successives, les connexions sont toujours dirigées des couches inférieures vers les couches supérieures et les neurones d'une même couche ne sont pas interconnectés. Un neurone ne transmet pas son état, qu'à un neurone est situé dans une couche postérieure à la sienne ([Hornik, 1989](#)).

3. Cluster Analysis

Il y a deux grandes approches en classification : la discrimination (classement) et la classification automatique (clustering), pour les méthodes type « clustering » qui est l'une des méthodes statistiques qui sont largement utilisées dans l'analyse de données. Il a été enfermé dans le cadre de l'apprentissage non supervisé, qui tente de parvenir des informations sans aucune connaissance antérieure, cependant ce n'est pas le cas pour l'apprentissage supervisé, la principale question est autour de laquelle se base le travail du Clustering, est de savoir d'imiter le mécanisme humaine d'apprentissage sans aucune information disponible auparavant, en utilisant des techniques qui permettent d'apprendre à partir certain nombre de données et de règles, selon certains caractéristiques sans aucune expertise ou intervention nécessaire. Ce processus nécessite certaines combinaisons ou traitement ou avec d'autres méthodes, en avant ou en post-processing, surtout pour une grande masse de données, afin de bien réaliser intégralement sa tâche de classification, L'ensemble des méthodes de traitement est généralement regroupé sous le terme de « fouille de données » ([Hartigans, 1975](#)).

3.1. Définition

Le clustering également connu sous le nom de segmentation qui est un regroupement en classes homogènes consiste à représenter un nuage des points d'un espace quelconque en un ensemble de groupes appelé « Cluster ». C'est un traitement sur un ensemble d'objets qui n'ont pas été ordonnés par un superviseur.

Généralement lié au domaine de l'analyse des données comme ACP (analyse linéaire en composantes principales) (Saporta, 1990 ; Lebart et al. 2000), ce type de méthodes vise à répondre au problème de : la diminution de la dimension de l'espace d'entrée, ou pour le groupement des objets en plusieurs catégories (clusters) non définies à l'avance. Parmi les techniques qu'on peut les trouver dans ce type de classification : les cartes auto-organisatrices de kohonen (kohonen, 1982), GMM . . .etc

Un «Cluster» c'est une collection d'objets qui sont «similaires» entre eux et qui sont «dissemblables » par rapport aux objets appartenant à d'autres groupes. Cette définition est clairement visible dans la figure ci-dessous.

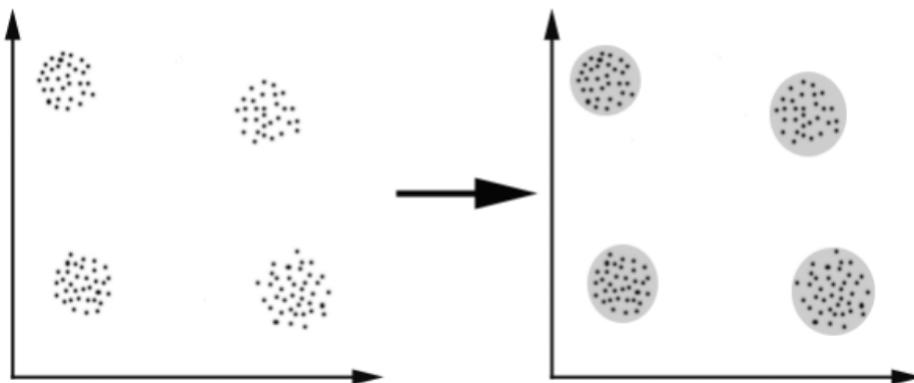


Fig. -8. Illustration de regroupement en clusters (in Koudri, 2011).

Dans ce cas, il est très facile pour une personne d'identifier 4 Clusters dans lesquels les données (nuage des points) peuvent être divisées (fig.8), les critères de similarité est la distance : deux ou plusieurs objets appartiennent au même cluster s'ils sont «proches», cela dépend d'une distance donnée (dans ce cas la distance géométrique).

Le clustering conceptuel est un autre type de rassemblement deux ou plusieurs objets appartiennent au même cluster si celui-ci définit un concept commun à tous les objets. Autrement, les objets sont regroupés en fonction de leur adéquation aux concepts descriptifs, et non en fonction d'une simple mesure de similitude.

3.2. Principe général

Contrairement à la classification (méthodes supervisées), nous n'avons aucune connaissance préalable des classes d'éléments prédéfinies. Par conséquent, les objets sont divisés en différents groupes (clusters) en fonction du calcul de la similitude entre les éléments. Bien que le but des méthodes de clustering soit de regrouper des éléments étroitement liés dans le même groupe, de sorte que deux éléments de données dans le même groupe soient aussi similaires que possible, tandis que deux éléments dans deux groupes différents sont aussi différents que possible (Harris & Racine, 1979). Sur le plan mathématique, on a un ensemble X de N données montrées chacune par leurs P attributs. Par conséquent, le clustering consiste la création consiste de créer une partition ou sa décomposition en sous-parties (clusters), par exemple:

- * Les données qui sont au même groupe se ressemblent,
- * Les données appartenant à deux groupes différents soient peu ressemblantes.

Nous utilisons fréquemment ce type de classification dans le traitement des images pour fixer les divers objets qu'elles contiennent (segmentation) : routes, villes, rues , des organes humaines (pour les images médicales) . . .

3.3. Les exigences de clustering

Les exigences essentielles soit, qu'un algorithme de clustering doit répondre sont les suivantes Evolutivité des clusters (in Koudri, 2011).

- traiter les différents types d'attributs.
- découvrir les clusters de forme arbitraire.
- exigences minimales pour la connaissance du domaine afin de déterminer les paramètres d'entrée.
- capacité de composer avec le bruit et les valeurs manquantes traitées les dimensionnalités.
- élevées. l'intelligibilité et la convivialité.

3.4. Les types de clustering

Il existe deux grands types du clustering :

A/ le clustering hiérarchique : d'agglomération («bottom-up»)

B/ le clustering non-hiérarchique : de division («top-down»)

Nous décomposons pour qu'un premier, l'ensemble d'individus en une arborescence de groupes. Dans le deuxième, on décompose l'ensemble d'individus en K groupes, les

algorithmes de ce type peuvent aussi être utilisés comme algorithmes de division dans le clustering hiérarchique (Fig. 9).

Mais dans certains ouvrages on classifie les types des algorithmes de clustering en quatre (4) groupes à cause des méthodes qui n'obéissent plus les normes du premier classement comme le cas de la règle « *Chaque objet doit appartenir à un seul groupe.* » Bien que la version floue l'ajustera et dans une certaine mesure permettra à un objet d'appartenir à plus d'une classe, les 4 types sont:

1. Clustering exclusive.
2. Overlapping clustering (fuzzy clustering).
3. Clustering hiérarchique.
4. Clustering probabiliste.

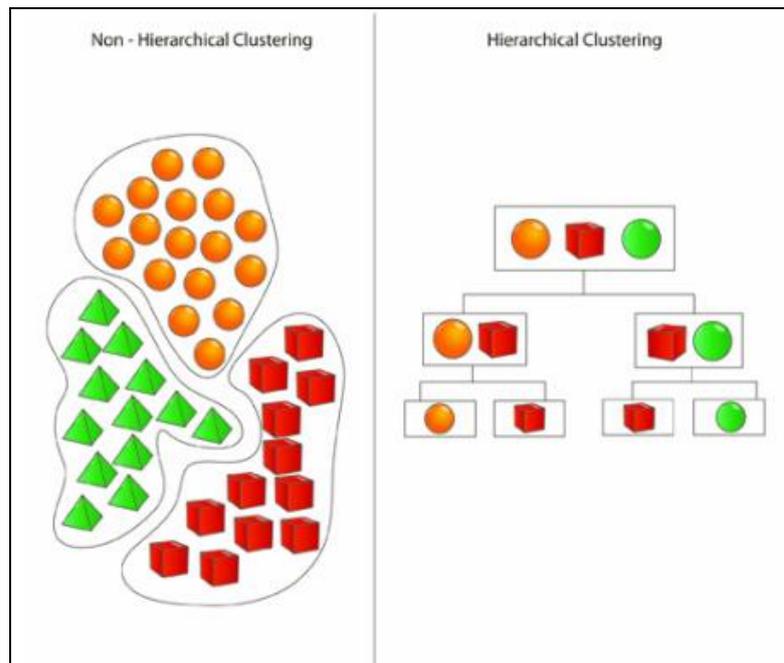


Fig. 9 -Deux types de clustering hiérarchique et non hiérarchique
(<https://www.jigsawacademy.com/cluster-analysis-for-business/>)

Dans le premier cas, les données sont regroupées de manière exclusive, donc si une donnée appartient à un cluster défini, elle ne peut pas être incluse dans un autre cluster. La figure ci-dessous montre un exemple simple où l'espacement des points est défini par des lignes droites sur un plan bidimensionnel.

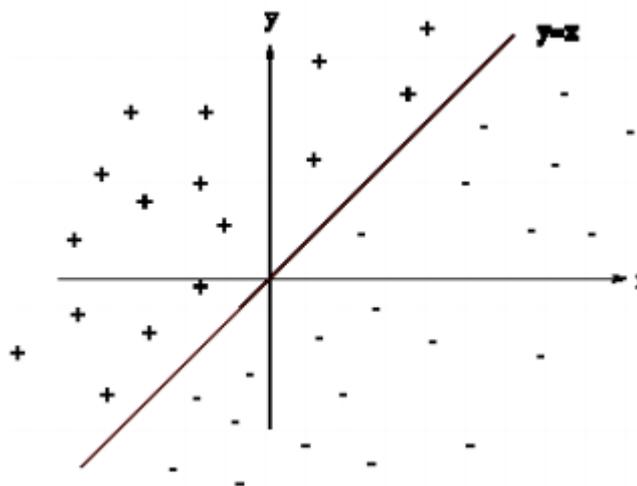


Fig.10 – Discrimination à deux classes, avec une séparatrice linéaire
(in Koudri, 2011).

Au contraire le deuxième type, le regroupement overlapping utilisé des ensembles flous aux données de cluster, de sorte que chaque point peut appartenir à deux ou plusieurs groupes avec différents degrés d'appartenance (Fig. 10). Dans ce cas, les données seront associées à une valeur d'une composition appropriée.

Comme mentionné précédemment un algorithme de clustering hiérarchique est basé sur l'union entre les deux plus proches clusters c'est-à-dire: consiste à trouver des clusters successifs utilisant des clusters qui sont auparavant établis. Pour la première condition, il faut mettre, au démarrage, chaque objet dans un cluster distinct et les fusionner en clusters successivement plus grand. Après quelques itérations, on peut atteindre le cluster final voulu regroupant tous les sous-clusters (sous-partitions). Enfin, la dernière méthode de regroupement consiste à utiliser une approche totalement probabiliste basée sur la probabilité d'appartenir approche.

3.5. Les algorithmes de clustering

Il existe dans la littérature plusieurs algorithmes de clustering, parmi lesquels :

1. K-means
2. Fuzzy C-means
3. Hierarchical clustering
4. Mixture of Gaussians (Expectation maximisation)

Chacun de ces algorithmes appartient à l'un des types de clustering répertoriés. A titre d'exemple, K-means est un algorithme de clustering exclusif, alors que Fuzzy C-means est un algorithme de Overlapping Clustering, tandis que clustering hiérarchique, il s'agit évidemment du troisième type de clustering, et enfin Mélange de Gaussien est un

algorithme de clustering probabiliste. Nous présentons brièvement les principes de ces méthodes de clustering :

3.5.1 K-means

L'algorithme k-means est développé par McQueen en 1967 (MacQueen, 1967), est l'un des algorithmes d'apprentissage non supervisés les plus simples, elle est appelée algorithme des centres mobiles (Benzécri, 1973 ; Celeux et al, 1989), qui attribue chaque point au centre (Centroid) dans le cluster le plus proche. La moyenne de tous les points du cluster est le centre, ses coordonnées sont la moyenne arithmétique de chaque dimension séparée de tous les points du cluster, c'est-à-dire que chaque cluster est représenté par son centre de gravité.

3.5.2 Méthode fuzzy C-means

Fuzzy C-means (FCM) est une méthode de clustering qui permet aux objets de données d'appartenir à deux ou plusieurs clusters. Cette méthode a été dérivée de l'algorithme c-means (Ball & Hall, 1967), qui est le même que à l'algorithme k-means précédemment décrit, elle a été développée par Dunn (Dunn, 1973) en 1973 et améliorée par Bezdek (Bezdek, 1981) en 1981, il a été largement utilisée dans la reconnaissance des formes.

3.5.3 Méthodes hiérarchiques

La base des méthodes hiérarchiques est donné par les auteurs (Johnson, 1967 ; Lance & Williams, 1967), Ce type de clustering consiste à effectuer une série de regroupements en Clusters de moins en moins en moins de clusters fins en agréant l'objet le plus proche (un seul élément) ou le groupe d'objets (une Cluster-partition-) à chaque étape. Cela nous donne une structure de cluster en forme arborescence (Celeux et al., 1989). Cette approche utilise la mesure de similarité pour refléter l'homogénéité ou l'hétérogénéité des classes le principe est très simple, au début pour chaque individu forme une classe, soit n classes, donc on cherche pour réduire ce nombre de classe $n_{\text{new}} < n$ itérativement de sorte que dans chaque étape on fusionne deux classes ensemble ou ajouter un nouveau élément à une classe. La valeur de dissimilarité est appelée indice d'agrégation. Qui commence dans la première itération faible, et croîtra d'itération en itération.

Parmi les algorithmes plus connus de ce type : La classification ascendante hiérarchique (CHA) où le mot ascendante est utilisé pour montrer qu'elle part d'une

situation dont tous les individus représentent des clusters à part entière, puis nous cherchons les rassembler en classes de plus en plus grandes. Aussi le qualificatif "hiérarchique" désigne le fait qu'elle produit une hiérarchie, (une amélioration a été proposée en 2002 par P. Bertrand , appelée Classification Ascendante 2-3 Hiérarchique).

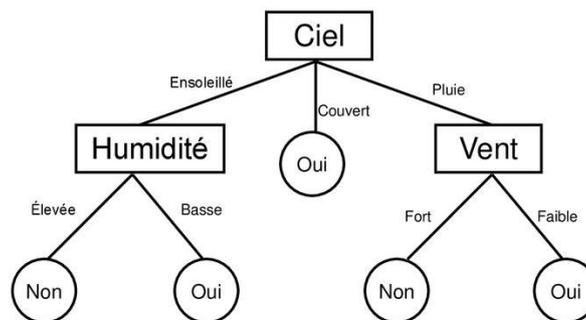
4. Random Forest (Forêts Aléatoires)

On présente d'abord l'arbre de décision classique et ses algorithmes de développement comme CART, la deuxième partie présentée par les méthodes d'ensemble et pour la troisième partie est concerné les forêts aléatoires (Random Forest).

4.1. Arbre de décision

4.1.1. Principe

Un arbre de décision est une méthode de de non seulement pour la prédiction et encore pour classification la plus populaire en apprentissage supervisé, elle est simple à utiliser et interpréter avec des performances très tolérables, elle donne des bonnes résultats satisfaisants pour des problèmes complexes (forêts aléatoires) (fig.11).



Classifier l'instance suivante:

<Ciel = Ensoleillé, Température = chaud, Humidité = élevé, Vent = fort>

4

Fig.11 - Exemple d'un arbre de décision (<https://slideplayer.fr/slide/14577667/>)

Il est largement utilisé dans de l'ingénierie, les domaines de statistiques, de la théorie de la décision, et aussi de l'apprentissage automatique (Simon, 2009). Bien que la généralisation de la fonction des méthodes de segmentation par les arbres soit relativement nouvelle, elle remonte au début des années 1990 mais ses origines sont beaucoup plus anciennes. La phase de construction qui vient d'être décrite peut être réalisée grâce à une variété d'algorithmes, et son dernier développement ne masquera certainement pas les origines antérieures.

Retracer ici quelques grandes étapes de cette évolution, pour s'intéresser par la suite plus spécialement à l'algorithme « *Random Forests* »

4.1.2. Partition de l'espace

Les méthodes de segmentation par arbres utilisent l'information qui sont disponibles concernant les individus non plus globalement, cependant de manière hiérarchisée. Les procédures de classement ont été basées sur la partition récursive de l'espace des attributs de manière à former en regard de la variable cible des entités d'homogénéité croissante. Cette homogénéisation est atteinte par un cheminement au travers d'une structure de décision hiérarchique. Chaque étape de ce approche a pris la forme d'un nœud de décision, qui se matérialisé par un test dont le résultat permettant de diviser le groupe d'individus I_0 ayant atteint ce nœud en une série de sous-groupes distincts I_1, I_2, \dots, I_n , eux-mêmes sujets a une éventuelle partition, et ainsi de suite. Ces estimateurs sont issus par les méthodes de segmentation récursive sont donc naturellement présentes sous forme des arbres, qui sont définis par la théorie mathématique ; "des graphes orientés acycliques possédant une racine unique" (in Brostaux, 2005).

On retrouve en groupement avec cette terminologie un nomenclature descriptif en analogie avec son homologue végétal. L'estimateur (point d'entrée), constitue par le premier test, qui forme ainsi la racine de l'arbre, pour lequel le test se dirige les individus vers les différentes branches selon les résultats trouvés, branches qui se subdivisent a leur tour grâce à d'autres tests, pour chacune des points de connexion entre plusieurs branches prene le nom «nœud », pour aboutir enfin aux nœuds terminaux qui sont appelés les feuilles, dans le cas des arbres de décision portent la prédiction finale (in Brostaux, 2005).

4.2. Random Forest (Forêts Aléatoires)

Random Forest (RF) en français s'appelle « les forêts aléatoires » qui introduits par Breiman en 2001, qui a donné une définition pour les forêts aléatoires comme une famille des méthodes d'ensemble (in Simon, 2009).

L'algorithme Random Forests (Breiman, 2001) est l'un des derniers résultats de la recherche consacrée à l'agrégation d'arbres randomisés. Assemblant les approches développées respectivement par (Breiman, 1997), il génère un jeu d'arbres doublement perturbés au moyen d'une randomisation opérée à la fois au niveau de l'échantillon d'apprentissage et des partitions internes. C'est une approche de l'ensemble qui permet de réaliser la haute exactitude sur la régression et la classification avec une réduction minimale des paramètres (Qing-You Zhang 2007). Pour chaque arbre du jeu

est ainsi fait au départ d'un sous-échantillon bootstrap du jeu d'apprentissage complet, de manière similaire aux techniques de bagging. Ensuite,

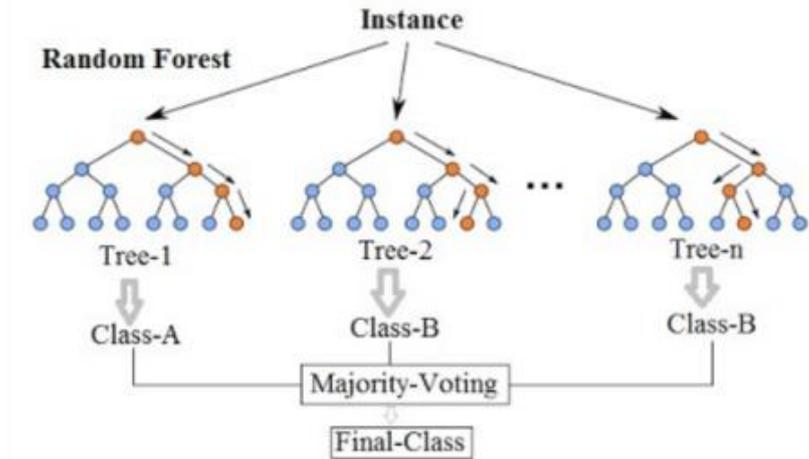


Fig.12 - Random Forest (Forêts Aléatoires) (https://f-origin.hypotheses.org/wp-content/blogs.dir/2973/files/2020/01/evs2020_r%C3%A9mi_ratajczak.pdf)

L'arbre est construit à l'aide de la méthodologie CART, La différence est qu'à chaque nœud, la meilleure sélection de partition basée sur l'indice de Gini n'est pas basée sur l'ensemble d'attributs M complet, néanmoins sur un sous-ensemble qui sélectionne de manière aléatoire au sein de celui-ci. La taille F de cette sélection est fixée précédemment à l'exécution de la procédure ($1 \leq F \leq M$) (Brostaux, 2005).

Par conséquent L'arbre est ainsi développé jusqu'a sa taille maximale, sans élagage. Pendant de la phase de prédiction les individus sont classés et est propagés dans chacun des arbres de la forêt et en fonction des règles CART a classifié. La prédiction globale de la forêt a été fournie par des votes à la majorité simple des attributions de classe des arbres individuels. Il est noté que les techniques de bagging constituent lors un cas particulier de l'algorithme Forest-R pour lequel $F = M$. Cet algorithme appartient à la famille plus large des forêts aléatoires, définie par (Breiman, 2001).

4.2.1. Définition d'une forêt aléatoire

Random forest est un classificateur qui en une collection de prédicateurs structures en arbres $[T(x, Q_k), k = 1,]$ ou les $[Q_k]$ sont des vecteur aléatoires des mêmes distributions et chaque arbre consiste par un vote unitaire de la plus populaire classe pour chaque entrée x .ä...

4.2.2. L'avantage de la méthode

Le essentiel avantage de cette structure est qu'elle permet d'éviter les dangers qui peuvent tomber sur-apprentissage pour toute méthode de prédiction est basée sur l'induction. C'est une méthode du non paramétrique haut dimensionnelle qui travaille bien sur grands nombres de variables (Qing-You, 2007). Breiman, 2001 démontre que lorsque le nombre d'arbres qui sont impliqués dans la forêt de prédiction augmente, le taux d'erreur en généralisation se converge vers une valeur limite, dont une borne supérieure peut être estimée sur base des caractéristiques intrinsèques de la forêt.

4.2.3. Fonction de la forêt aléatoire

C'est le degré de confiance du classement qui a été établi par les arbres de cette forêt sur la population mesuré par la différence de probabilité entre la meilleure classe erronée et la prédiction de la classe correcte Y , on peut définir la valeur de la prédiction d'un jeu d'arbres par l'espérance mathématique de cette fonction La dépendance entre arbres d'une forêt est quand a elle mesurée par la corrélation entre leur fonction marginale brute, évaluée pour des valeurs de paramètres fixées et distinctes (in Brostaux, 2005).

4.2.4. Les techniques d'agrégation sur bootstrap ou bagging

Les méthodes d'agrégation sur bootstrap ou bagging ont été développées par Breiman, 1996, et ont été marqués par une étape importante dans ce domaine, mais des travaux antérieurs abordent des problèmes similaires, liés à la synthèse des résultats d'arbres multiples.

Le principe du bagging est simple : pour construire un estimateur synthétique représentatif d'une population x , la solution est d'utiliser k échantillons indépendants E_1, E_2, \dots, E_k issus de cette population pour entraîner k estimateurs. Ensuite, la prédiction globale de cette séquence d'estimateur est obtenue par l'espérance mathématique de la variable cible, et estimée par la moyenne des prédictions qui sont observées dans le cas d'une variable cible numérique (régression), ou par vote à la majorité simple pour les problèmes de classification. Cependant, en fait, nous n'avons pas une population x , mais seulement un échantillon E , qui a une taille de n . Au fur et à mesure que le nombre d'employés diminue et que l'instabilité augmente, il n'est pas judicieux de diviser ces derniers en k sous-ensembles indépendants.

Le processus d'échantillonnage initial peut toutefois être imité par une série d'échantillons bootstrap de E , c'est à dire par la répétition de k tirages aléatoires avec la remise d'effectifs égaux à n dans l'échantillon d'apprentissage. Ces échantillons perturbés sont utilisés comme base pour construire k estimateurs, dont les résultats sont synthétisés de la manière décrite ci-dessus, de sorte que le nom d'agrégation bootstrap ou ensachage donne ces méthodes. En outre, l'utilisation du bootstrap autorise une nouvelle estimation de l'erreur réelle de prédiction. Cela ne nécessite plus l'utilisation ni d'un échantillon test, ni de la validation croisée, mais recourt aux individus non repris dans le tirage avec remise. Cette estimation (out of bag "OOB") est la même que l'estimation précédente, non biaisée et plus précise (in Brostaux, 2005).

4.2.5. Estimation out of bag "OOB"

Parallèlement aux excellents résultats dans la phase de prédiction, la structure de l'algorithme Random Forest lui permet également de fournir des renseignements complémentaires concernant l'estimateur qu'il construit. L'utilisation d'échantillons bootstrap permet en particulier le calcul du taux d'erreur out-of-bag (OOB), qui fournit une estimation non biaisée du taux d'erreur généralisé sans avoir recours à un échantillon test supplémentaire. Deux autres informations peuvent aussi être calculées sur la demande, la mesure de l'importance des différents attributs et une mesure de proximité des individus classes (Brostaux, 2005 ; Qing-You, 2007).

4.2.6. Matrice de confusion

La matrice de la confusion créée pendant l'apprentissage utilise la (out of bag OOB). Ce dernier fournit des informations détaillées sur la manière dont le modèle classe les lignes de données pour chaque catégorie de la cible variable a une ligne et colonne. La catégorie dans la première colonne est la catégorie réelle de la variable cible. Les catégories affichées en haut du tableau sont des catégories prédites. Le nombre dans la cellule correspond à la catégorie réelle de la ligne de pondération de la ligne de données avec pondération et à la catégorie prévue de la colonne. Voici un exemple de matrice de confusion.

Confusion Matrix			
Actual : Category : ----- :	----Predicted Category----		
	1	2	3
1 :	17	1	4
2 :	1	13	2
3 :	3	1	26

Fig.13 - Matrice de confusion (in Saadi, 2009).

Les nombres trouvés dans les cellules diagonales représentent les poids pour les cas correctement classés où la catégorie réelle égale à la prédite. Les cellules qui sont hors de la diagonale, elles sont les lignes mal classées.

Par exemple : La matrice de la confusion (Fig.13) a montré, qu'il y a un total de $(17+1+4)=22$ cas de cible catégorie (1. 17) ont été prédits correctement pour être catégorie (1, 1) est prédit pour être catégorie 2, et 4 ont été prédits pour être catégorie 3. "La matrice de confusion" donnera l'indication pour le modèle comment il travaillera avec les données différent que les données d'apprentissages.

4.3. La dimension de modèle

La dimension du modèle représente le nombre d'arbres dans la forêt, qui spécifie le nombre d'arbres qui seront construits dans la forêt des arbres de décision. La méthode «forêt aléatoire» donne un graphique qui représente la taille du modèle (le nombre d'arbres dans la forêt) et le taux d'erreur de classification. Le but de cette figure est de mettre en évidence la relation entre le nombre d'arbres et le taux d'erreur. La sélection du nombre d'arbres liés au taux d'erreur. Par conséquent, la meilleure forêt fournit le plus petit taux d'erreur. En général, plus la forêt de l'arbre de décision est grande, plus la prédiction est précise.

5. Algorithmes Génétiques

5.1. Concepts et principe des algorithmes génétiques

Dans son usage le plus général, les algorithmes génétiques font référence à une famille de modèles de calcul inspiré par l'évolution. Ces algorithmes codent une solution potentielle à un problème spécifique sur une simple structure de données semblable à un chromosome et appliquer des opérateurs de recombinaison à ces structures afin de

préserver les informations critiques. Les algorithmes génétiques sont souvent considérés comme optimiseurs de fonction, bien que l'éventail des problèmes auxquels les algorithmes génétiques appliqué est assez large (Whitley, 1994). Depuis que le premier travail de John Holland sur les systèmes adaptatifs remonte à 1962, les algorithmes génétiques ont une longue histoire in (in Hamida, 2015).

L'algorithme génétique recherche l'extremum de la fonction définie dans l'espace de données. Pour l'utiliser, vous devez avoir les cinq éléments suivants (Whitley, 1994):

1. Le principe de codage de l'élément de population.
2. Un mécanisme de génération de la population initiale.
3. Une fonction à optimiser est appelée (fitness), c'est une fonction de l'évaluation des individus.
4. Des opérateurs permettent de varier la population au cours des reproductions et de l'explorer d'espace d'état. L'opérateur de croisement reconstitue les gènes d'individus qui existent dans la population, l'opérateur de mutation est de garantir l'exploration de l'espace d'état.
5. Des paramètres de dimensionnement qui sont la taille de la population, le nombre total de générations ou critère d'arrêt, et les probabilités d'application des opérateurs de croisement et de mutation.

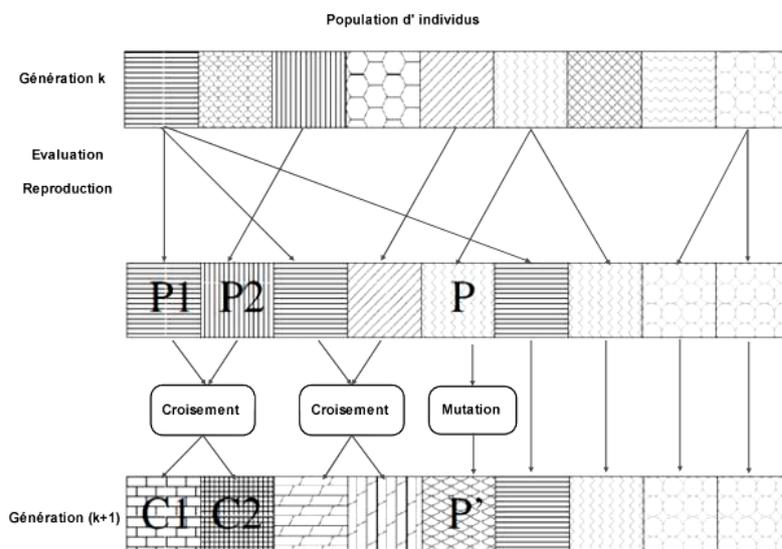


Fig.14 - Principe des algorithmes génétiques (in Hamida, 2015).

Afin de trouver la meilleure solution au problème à l'aide d'algorithmes génétiques, nous générons d'abord de manière aléatoire un ensemble de solutions (un ensemble d'individus). L'évolution d'une génération à l'autre utilise trois opérations (sélection, croisement et mutation) appliquées à tous les éléments de populations. Les

parents P1 et P2 sont sélectionnés en fonction de leur adaptation. L'opérateur de croisement a été appliqué avec la probabilité P_c et génère une paire de E1 et E2. Après, l'opérateur de mutation est appliqué à la progéniture avec la probabilité P_m , et les individus mutants E1 'et E2' sont générés et insérés dans la nouvelle population. Parmi les critères d'arrêt qui peuvent être retenus pour l'algorithme, on peut citer le nombre maximum de générations fixées.

5.2. La mise en œuvre d'un algorithme génétique

La mise en œuvre d'un AG est réalisée suivant les étapes:

1. Création d'une population initiale.
2. Evaluation des individus de la population.
3. Sélection des meilleurs individus.
4. Reproduction (Croisement et mutation).
5. Formation d'une nouvelle génération.

La figure suivante montre l'organigramme de fonctionnement d'un algorithme génétique.

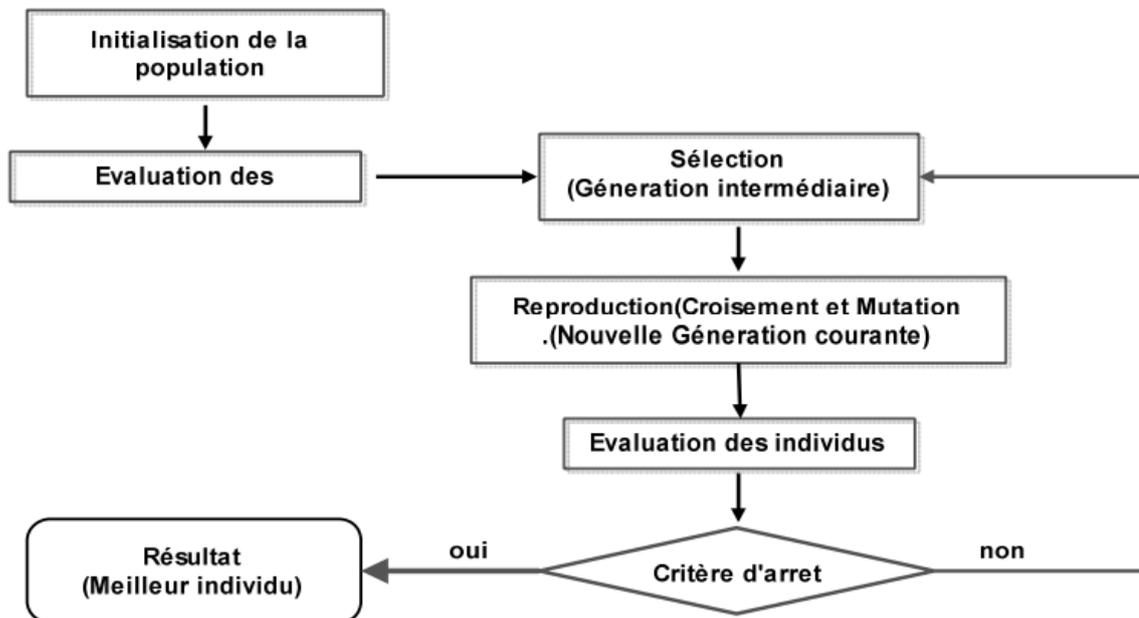


Fig. 15 - Organigramme de l'algorithme génétique (in Hamida, 2015).

5.3. Vocabulaires des algorithmes génétiques

Dans cette section nous introduisons quelques vocabulaires utilisés dans la mise en œuvre des algorithmes génétiques (Durand, 2004)

- Individu : représenté par un chromosome (génome).
- Un chromosome est une chaîne de gènes.
- Génotype : l'ensemble des gènes représentés par un chromosome.

- Phénotype : l'ensemble des valeurs observables prises par chaque gène.
- Fonction d'adaptation : fitness.
- Opération de reproduction :
 - Le croisement.
 - La mutation.
- Génération : l'ensemble de la population à un moment donné du processus.

5.4. Le codage

Le codage est la modélisation de la solution à un problème donné sous la forme d'une série de caractères appelés chromosomes, où chaque caractère (également appelé gène) représente une variable ou une partie du problème. La principale tâche est de sélectionner un contenu génétique permettant de décrire le problème et de respecter ses contraintes. Il utilise particulièrement deux types de codage qui sont le codage binaire et le codage réel (in Hamida, 2015) :

5.4.1. Codage binaire

Dans ce type de codage, les chromosomes représentent que des séquences de 0 et 1. Le codage binaire est indépendant des opérateurs génétiques (mutation et croisement), car ces derniers ne nécessitent aucune spécification. En fait, toute manipulation des chromosomes produira un nouveau chromosome efficace. Cependant, ce codage n'est pas obligatoirement bon et parfois les solutions de codage de cette manière sont très difficiles ou très lourds. En outre, dans certains cas, la taille de mémoire requise peut devenir trop exagérée.

5.4.2. Codage réel

Le codage est pour chaque chromosome un vecteur avec ses composantes qui sont les paramètres du processus d'optimisation. Par exemple, si nous recherchons la fonction optimale de n variables, nous pouvons simplement utiliser un chromosome contenant n variables pour ce type de codage. Comme il n'y a pas d'étapes suivantes, le processus d'évaluation des chromosomes sera plus rapide à coder et à transcoder (du réel vers le binaire et vice-versa).

5.5. La population initiale

Le choix de l'individu initial détermine fortement la vitesse de l'algorithme. Afin de générer la population initiale, nous avons deux possibilités. Si aucune information sur l'emplacement de la solution n'est disponible, l'objectif est de couvrir autant que possible

l'espace d'états. Par conséquent, la génération aléatoire est générée par des tirages uniformément chaque gène sur le chromosome

Si l'utilisateur connaît le sous-domaine de l'espace d'états dans lequel la solution est trouvée, la population initiale est générée dans ce sous-espace. Pour les générations générées aléatoirement, il existe deux possibilités. Vous pouvez savoir à l'avance si un certain point satisfait aux contraintes du problème d'optimisation, puis générer de manière aléatoire une population dans le domaine autorisé. Soit il est impossible de le savoir, soit une pénalisation est ajoutée à la fonction objective pour s'assurer que les contraintes sont respectées, et elles sont généralement uniformément réparties dans tout l'espace d'états.

5.6. Fonction d'évaluation (fitness)

L'évaluation de l'adaptabilité de chaque individu à l'environnement est réalisée à travers des fonctions adaptatives (fitness). Cette fonction est attribuée à chaque individu une valeur représentant son niveau d'adaptation. La fonction adaptative directement peut affecter la qualité et le temps d'exécution des résultats obtenus par l'algorithme génétique. Cela dépend des critères à minimiser. Du fait de l'analogie avec la théorie de l'évolution (la survie de l'individu la mieux adaptée à son environnement), l'algorithme génétique est naturellement basé sur la maximisation.

5.7. Les opérateurs de base d'un AG

Les opérateurs sont fondamentaux pour implanter le processus reproductif caractéristique d'un AG. De nombreux opérateurs génétiques existent ainsi que différentes stratégies d'implantation. Par contre, dans ce paragraphe nous n'abordons que les opérateurs les plus simples et les plus utilisés à savoir : l'opérateur de sélection, de croisement et de mutation, pour ne pas complexifier la méthode et perdre le sens physique des paramètres de réglage.

5.7.1. Le croisement

La naissance d'un nouvel individu, nécessite la prise aléatoire d'une partie des gènes de chacun des deux parents. Ce phénomène, issu de la nature est appelé croisement (crossover). Il s'agit d'un processus essentiel pour explorer l'espace des solutions possibles. Une fois la sélection terminée, les individus sont aléatoirement répartis en couples. Les chromosomes parents sont alors copiés et recombinaison afin de produire chacun deux descendants ayant des caractéristiques issues des deux parents. Dans le but de garder quelques individus parents dans la prochaine population, on associe à

l'algorithme génétique une probabilité de croisement, qui permet de décider si les parents seront croisés entre eux ou s'ils seront tout simplement copiés dans la population suivante. Il existe plusieurs types de croisement parmi lesquels on trouve : le croisement en 1 point, le croisement en deux points et le croisement en N points, ces types sont résumés dans la figure suivante :

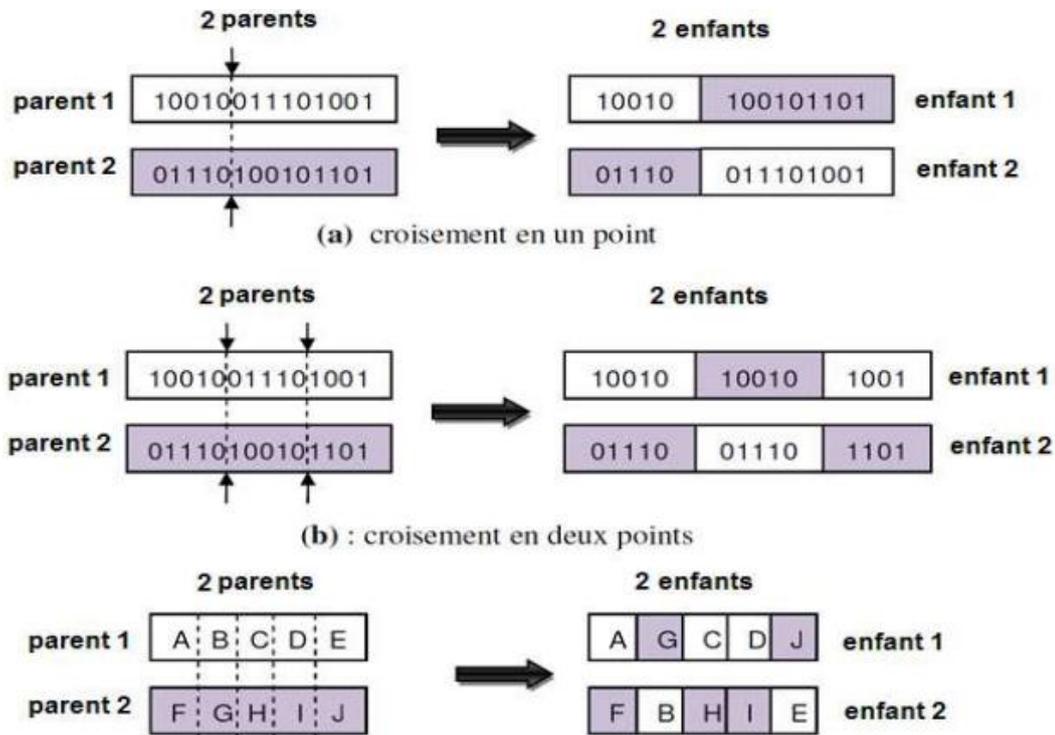


Fig. 16 - Exemples d'opération de croisement (in Hamida, 2015).

5.7.2. La mutation

L'opérateur de mutation, en générant de nouveaux gènes, a pour rôle de permettre d'explorer la totalité (en théorie) de l'espace d'état, ce qui correspond à la propriété d' périodicité de parcours d'espace, essentielle aux AG pour leurs propriétés de convergence. En effet les preuves théoriques de convergence des AG peuvent fonctionner sans croisement, mais pas sans mutation.

L'opérateur de mutation fonctionne comme suit. Pour les problèmes discrets, un gène du chromosome est tiré aléatoirement et sa valeur est remplacée par une des autres valeurs possibles (tirée aléatoirement elle aussi). Dans le cas des problèmes continus, le gène est également tiré aléatoirement, et remplacé par une valeur aléatoire du domaine d'extension des gènes (espace d'état). La figure suivante montre un exemple de mutation.

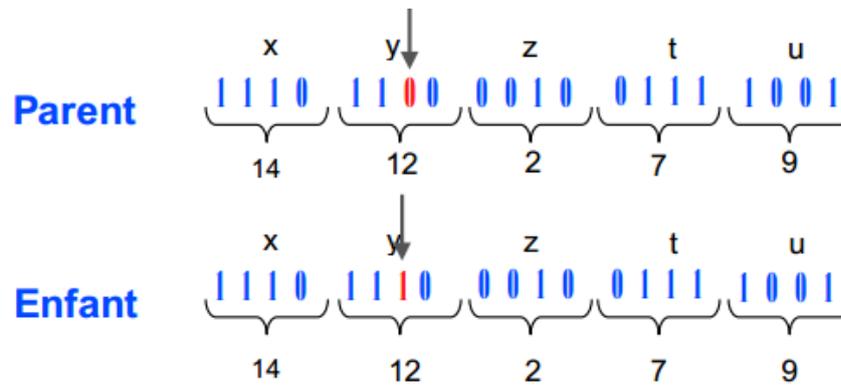


Fig. 17 - Exemple de mutation (in Hamida, 2015).

5.7.3. La sélection

L'opérateur de sélection est chargé de "favoriser" les meilleurs individus. Plus formellement, l'opérateur de sélection va générer à partir de la population courante une nouvelle population par copie des individus choisis de la population courante. La copie des chaînes s'effectue en fonction des valeurs de la fonction d'adaptation. Ce procédé permet de donner aux meilleures chaînes, une probabilité élevée de contribuer à la génération suivante. Cet opérateur est bien entendu une version artificielle de la sélection naturelle, la survie darwinienne des chaînes les plus adaptées. Plusieurs stratégies sont possibles pour effectuer une telle sélection parmi lesquelles nous abordons :

La sélection par classement : elle consiste à ranger les individus de la population dans un ordre croissant (ou décroissant selon l'objectif) et à retenir un nombre fixé de génotypes. Ainsi, seuls les individus les plus forts sont conservés. L'inconvénient majeur de cette méthode est la convergence prématurée de l'algorithme génétique.

La sélection par tournoi : le tournoi le plus simple consiste à choisir aléatoirement un nombre k d'individus dans la population et à sélectionner celui qui a la meilleure performance. Les individus qui participent à un tournoi sont remis ou sont retirés de la population, selon le choix de l'utilisateur. Avec le tournoi binaire, sur deux individus en compétition, le meilleur gagne avec une probabilité $p \in [0, 5; 1]$.

5.8. La convergence d'un AG

Le cycle de génération et de sélection de population est répété jusqu'à ce qu'un critère d'arrêt soit satisfait ; ce critère peut être notamment un nombre maximum de générations, un temps maximal de calcul, une valeur de fitness minimale, ou/et

une convergence vers une solution satisfaisante. Les valeurs de tels paramètres dépendent fortement de la problématique étudiée. Ainsi il n'existe pas de paramètres qui soient adaptés à la résolution de tous les problèmes qui peuvent être posés à un algorithme génétique. Cependant, certaines valeurs sont souvent utilisées et peuvent être de bons points de départ pour démarrer une recherche de solutions à l'aide d'un AG.

- La probabilité de croisement est choisie dans l'intervalle [0.7, 0.99].
- La probabilité de mutation est choisie dans l'intervalle [0.001, 0.01].

Les algorithmes génétiques constituent une classe de méthodes approchées ou métaheuristiques. Cette classe est adaptable à un très grand nombre de problèmes d'optimisation combinatoire (Saad, 2006 ; Tangour, 2008). Les AGs ont révélé leur grande efficacité pour fournir des solutions approchées de bonne qualité pour un grand nombre de problèmes d'optimisation classiques et d'applications réelles de grande taille.

On retrouve les algorithmes génétiques dans différents domaines d'applications, notamment en Bioinformatique, Phylogénie , Économie, Sciences numériques, Robotiques, Industries, Mathématique, Physique et Chimie, et même l'ingénierie de réservoir à une large utilisation pour trouver des solutions les problèmes de réservoirs (Irani & Nasimi, 2011; Kaydani , 2011 ; Bagheripour & Asoodeh, 2012 ; Yongchao et al., 2014).

6. Support Vector Machines (SVM)

Les machines à vecteurs de soutien ou séparateurs (SVM) sont un groupe de techniques d'apprentissage supervisé visant à résoudre les problèmes de discrimination. SVM est une généralisation des classificateurs linéaires. Les machines à vecteurs de support ont été développées dans les années 1990 sur la base des considérations théoriques de Vladimir Vapnik sur le développement de l'apprentissage de la théorie statistique: théorie de Vapnik-Chervonenkis (Vapnik, 1995). SVM est rapidement adopté en raison de sa capacité à traiter le big data, un petit nombre d'hyperparamètres, ses garanties théoriques et ses bons résultats en pratique. Le SVM a été appliqué dans de nombreux domaines (bioinformatique, recherche d'information, vision par ordinateur, finance et ingénierie des réservoirs ...). Selon les données, les performances.

6.1. Mise en oeuvre des SVM

La réalisation du programme d'apprentissage par SVM consiste essentiellement à résoudre un problème d'optimisation, qui implique le système de résolution de programmation quadratique dans l'espace des dimensions du résultat. C'est pour cette raison ces programmes utilisent des méthodes spéciales pour atteindre efficacement cet objectif, il n'est donc pas recommandé d'essayer de créer de tels programmes vous-même. Définissez les paramètres de ces fonctions). Ces choix sont généralement faits par des techniques de validation croisée, dans lesquelles la performance du système est évaluée par une évaluation empirique d'exemples non utilisés pendant la formation. L'idée est de trouver les paramètres qui permettront d'obtenir les meilleures performances.

6.2. Support Vector Machines

SVM est un algorithme d'apprentissage automatique supervisé qui peut être utilisé pour la classification et la régression. SVM est le plus couramment utilisé dans les situations de classification. SVM est basé sur l'idée de trouver l'hyperplan qui partage le mieux de l'ensemble de données en deux parties.

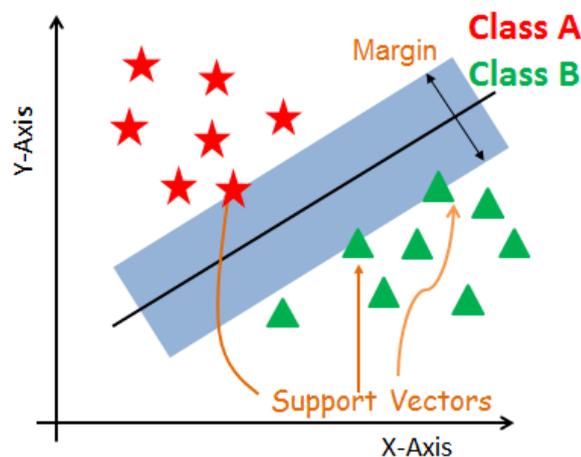


Fig. 18– Schéma du vecteurs de support

<https://www.datacamp.com/community/tutorials/svm-classification-scikit-learn-python>

Le vecteur de support est le point de données le plus proche de l'hyperplan (voir figure ci-dessus). Le point dans l'ensemble de données (s'il est supprimé) modifiera la position de l'hyperplan fractionné. Par conséquent, ils peuvent être considérés comme la partie la plus importante de l'ensemble de données.

6.3. Qu'est-ce qu'un hyperplan?

À titre d'exemple simple, pour une tâche de classification avec seulement deux fonctions (comme illustré dans la figure ci-dessus), on peut comprendre un hyperplan

comme une ligne pour séparer et classer linéairement un ensemble de données. Plus nos points de données sont éloignés de l'hyperplan, plus nous sommes confiants dans leur classement. Par conséquent, l'objectif est de garder nos points de données aussi loin que possible de l'hyperplan tout pour rester en sécurité. Mais par conséquent, lors de l'ajout de nouvelles données de test, quel que soit le côté de l'hyperplan sur lequel elles se trouvent, il détermine la catégorie que nous lui attribuons.

6.4. Comment trouver le bon hyperplan?

La distance entre l'hyperplan et le point de données le plus proche (l'un d'entre eux) est appelée marge. L'objectif est de sélectionner un hyperplan avec la plus grande marge possible entre l'hyperplan et n'importe quel point de la formation, ce qui augmente les chances de classification correcte des nouvelles données.

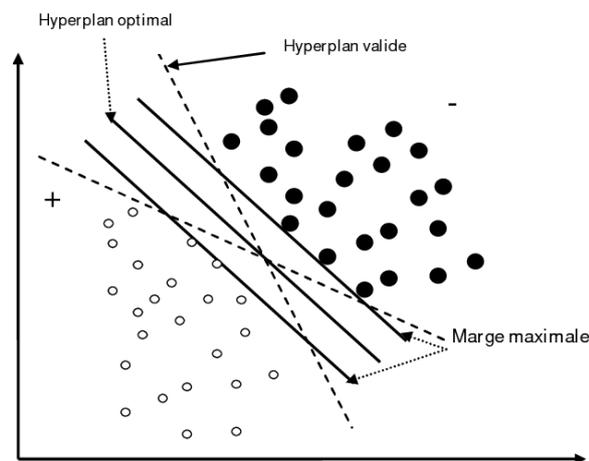


Fig. 19 - Le bon hyperplan optimal (in Djeffal, 2012).

6.5. Avantages et les inconvénients

6.5.1. Les avantages

- Une grande précision de prédiction.
- Une bonne Fonctionnement sur des petits data sets.
- Plus efficace parce que, ils utilisent un sous-ensemble de points d'entraînement.

5.6.2. Les inconvénients

- Le temps d'entraînement avec les SVM peut être long, alors il ne convient pas à des jeux de données plus volumineux,
- Il est moins efficace si les données contiennent du bruit et beaucoup des outliers.

6.7. Utilisations SVM

Le SVM il a une large utilisation surtout pour résoudre les problèmes de classification de texte tels que l'attribution de catégorie, la détection de spam et ainsi l'analyse des sentiments. Il est couramment utilisé pour la reconnaissance d'image, en particulièrement la reconnaissance de forme et la classification de couleur. Alors le SVM joue un rôle très essentiel dans de nombreux domaines (in Djeffal, 2012).

6.8. Type de noyau

Le produit scalaire entre les vecteurs individus tient une place importante dans les calculs (formulation duale), et elles peuvent tirer d'avantage des fonctions « noyau »

- Le noyau linéaire : $K(x, x_0) = \langle x, x_0 \rangle$ (correspondant aux SVM linéaires)
- Le noyau Gaussien radial (Gaussian RBF) $K(x, x_0) = \exp(-\sigma \|x - x_0\|^2)$
- Les noyaux polynomiaux $K(x, x_0) = (\alpha + \beta \langle x, x_0 \rangle)^\delta$
- Le noyau radial de Laplace (Laplace RBF) $K(x, x_0) = \exp(-\sigma \|x - x_0\|)$
- Le noyau tangente hyperbolique (sygmoïde) $K(x, x_0) = \tanh(\alpha + \beta \langle x, x_0 \rangle)$

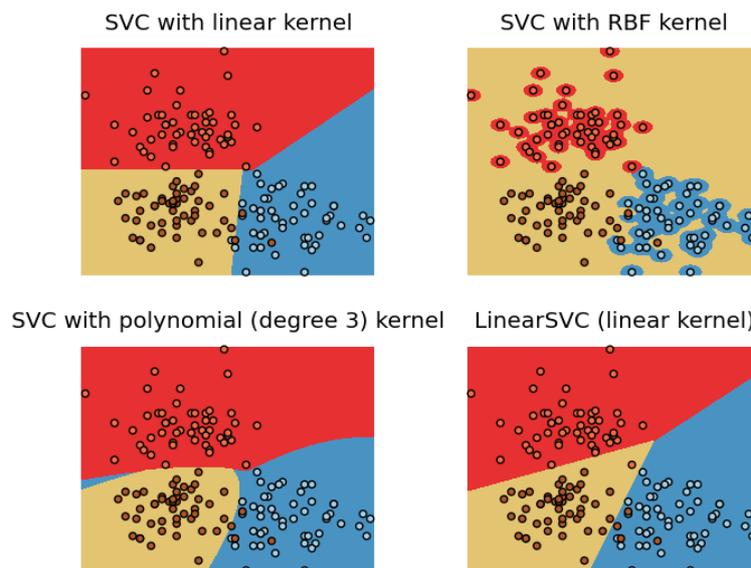


Fig. 20 - Types de noyau (kernel) de SVM

(<https://medium.com/all-things-ai/in-depth-parameter-tuning-for-svc-758215394769>)

6.9. Applications de SVM dans la vie

Les SVM sont dépendants des algorithmes d'apprentissage supervisé. Le but de l'utilisation de SVM est de classer correctement les données invisibles. Les SVM ont une large application dans plusieurs domaines. Certaines applications courantes de SVM sont:

6.9.1. Détection des visages

SVM divise les parties de l'image en deux parties, (visage et en non-visage) et encore créer une bordure carrée autour du visage.

a) Catégorisation du texte et de l'hypertexte

SVM prend en charge la classification textuelle et hypertexte des modèles d'induction et de transduction. Ils utilisent les données de formation afin de classer les documents en différentes classes. Il est classé en fonction des scores générés puis compare avec la valeur de seuil.

b) Classification des images

SVM donne une meilleure précision dans la recherche de classification des images. Il offre une bonne précision comparé avec les techniques traditionnelles.

c) Bioinformatique

Dans le domaine de la bioinformatique, on utilise les techniques de SVM pour pouvoir identifier la classification des gènes, des patients sur la base de gènes et d'autres problèmes biologiques.

d) Détection de repli protéique et d'homologie à distance

En appliquant SVM, on peut détecter d'homologie à distance de protéines.

e) Reconnaissance de l'écriture manuscrite

SVM ont été utilisés pour reconnaître les caractères manuscrits qui sont largement utilisés.

f) Détection des outliers

Un des points clefs de la robustesse de la marge adaptative est que l'on a la possibilité de détecter les points aberrants et donc de les retirer du jeu de données. Un exemple en deux dimensions de la procédure de détection de ces points aberrants est donné sur la figure 20. La ligne continue représente l'hyperplan optimal, elle est encadrée de 2 lignes pointillées qui représentent les limites de la marge d'insensibilité ". Les points positionnés à l'extérieur de cette marge sont appelés les outliers-SVs et les points positionnés sur les limites de cette marge sont appelés les margin-SVs.

Conclusion

Dans les dernières années, les méthodes de l'intelligence artificielle, ont le potentiel de résolution des problèmes en géoscience et l'industrie pétrolière, en particulier en cas de disponibilité limitée ou de manque de données. Alors l'application de l'IA est devenue indispensable dans l'ingénierie, y compris la géoscience et l'ingénierie pétrolière.

A titre d'exemple on peut citer quelques travaux d'application des différentes techniques de l'intelligence artificielle pour résoudre les problèmes dans le domaine pétrolier ; les données sismiques (Diersen, et al., 2011), la géochimie (Sun et al., 2009), le forage (Kannan et al., 2014), le well testing (Adibifard et al., 2014) et les diagraphies (Salehi et al., 2017), pour générer des mesures de porosité et de perméabilité à partir de l'enregistrement de diagraphies (Jamialahmadi & Javadpour, 2000; Lim, 2005; Mohaghegh, 2005). Prédiction de la porosité de fracturation (Zerrouki, 2014), Estimation du carbone organique totale (Handhal, 2020).

Chapitre II

Géologie Régionale et Locale de la zone d'étude

Chapitre II : Géologie régionale et locale de la zone d'étude

Introduction

Le bassin de Berkine a fait l'objet de plusieurs découvertes et recherches sur le plan géologie, en raison de sa richesse en d'hydrocarbures dans les différentes phases de l'exploration, de caractérisation et de développement (Boudjema, 1987; Echikh et al., 1998; Rossi et al., 2002; Gauthier, 2003; Ouadfeul et al., 2011; Kadi, et al 2013; Souadnia et al 2017; Baouche et al., 2020; Belhouchet et al., 2020; Ali-Khoudja et al., 2020). Le bassin intracratonique de Berkine couvre une superficie de 102 395 km² à l'est de l'erg (W.E.C 2007). Il est limité au nord par la flexion saharienne et les unités allochtone des montagnes du Sahara et de l'Atlas (Turner et al., 2001), et du nord du bassin d'Illizi au sud, par les frontières tuniso-libyennes à l'est et le haut d'Amguid El Biod – Hassi Messaoud à l'ouest (Fig. 21).

Le synclinal de Berkine a été structuré pendant les périodes de paléozoïque et du mésozoïque, préservant une couverture sédimentaire de 7 000 m d'épaisseur (Yahi, 1999 ; Galeazzi et al. 2010). Ces sédiments sont séparés par une discordance régionale: taconienne, calédonienne, hercynienne et autrichienne.

1. Géologie Régionale du Bassin de Berkine

1.1. Situation géographique

Le bassin de Berkine (anciennement Ghadames) est situé dans la partie orientale, du Sahara de l'Algérie s'appelé (l'Erg oriental) avec 29 ° à 34 ° de latitude nord et 5 ° à 10 ° de longitude est (figure 21), les limites géographiques du bassin de Berkine sont :

- Le bassin de Syrte entre la Tunisie et la Libye à l'Est, avec une série de failles de direction NW-SE.
- Le bassin d'Illizi au sud.
- La mole d'Amghuid El Biod Hassi Messaoud à l'ouest.
- Le bourrelet d'Ain-Roumana et la voute de Dahar au nord.

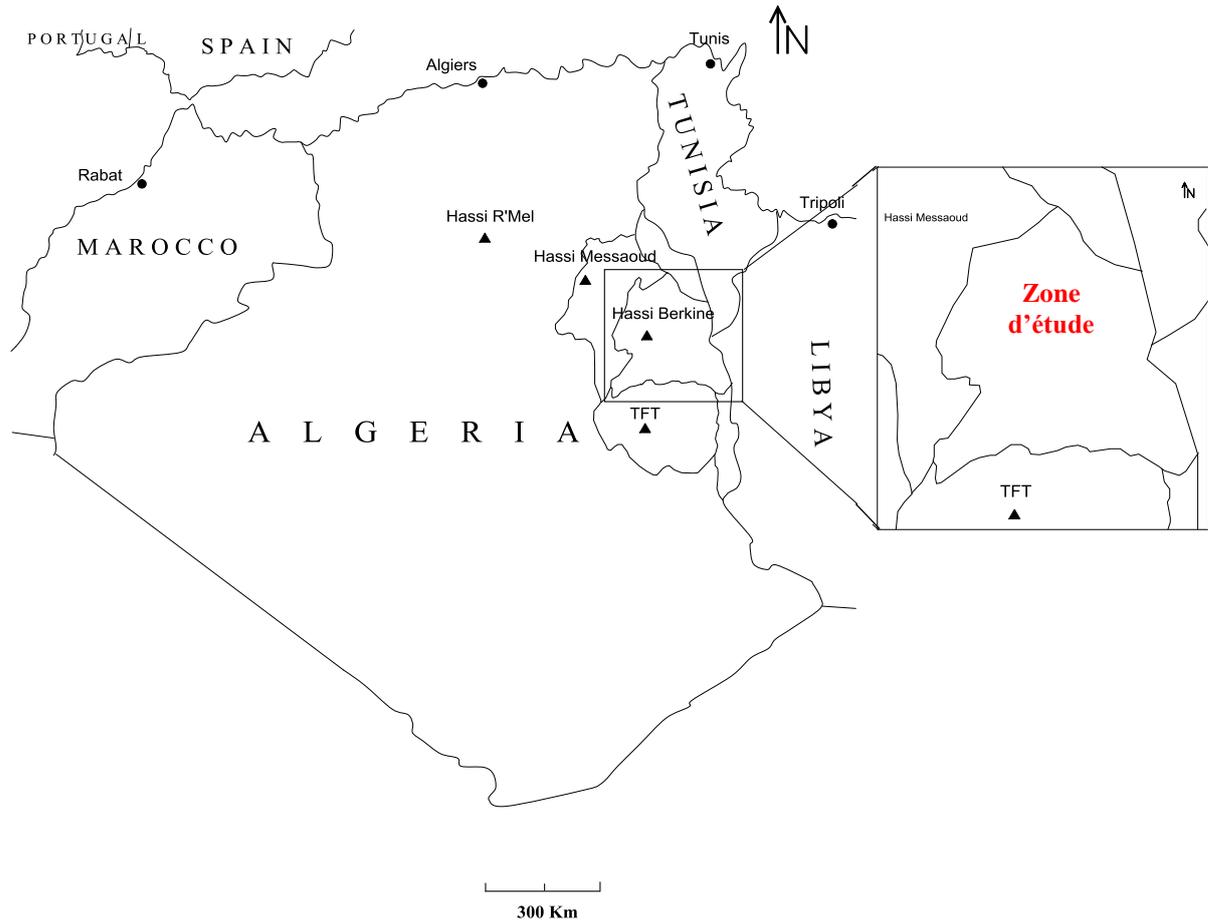


Fig. 21 -Situation géographique du bassin de Berkine.

1.2. Cadre géologique régionale

Le bassin de Berkine est classé comme un type intracratonique (W.E.C.2007), il se situe dans la partie Nord Est de la plate-forme saharienne. Il est défini comme un bassin de plate-forme à substratum infracambrien. Ce bassin occupe une superficie totale environ de 300.000 Km², mais uniquement 103.000 Km² sont appartenent l'Algérie les autres sont partagés entre la Tunisie et la Lybie. Le socle a été rencontré à une profondeur de 7000 m, et il est constitué par des roches cristallophylliennes, volcaniques et métamorphiques (Askri, 1995).

Sur le socle cristallophyllien, il repose en discordance une pile sédimentaire comprenant des termes, allant du cambrien jusqu'au l'actuel.

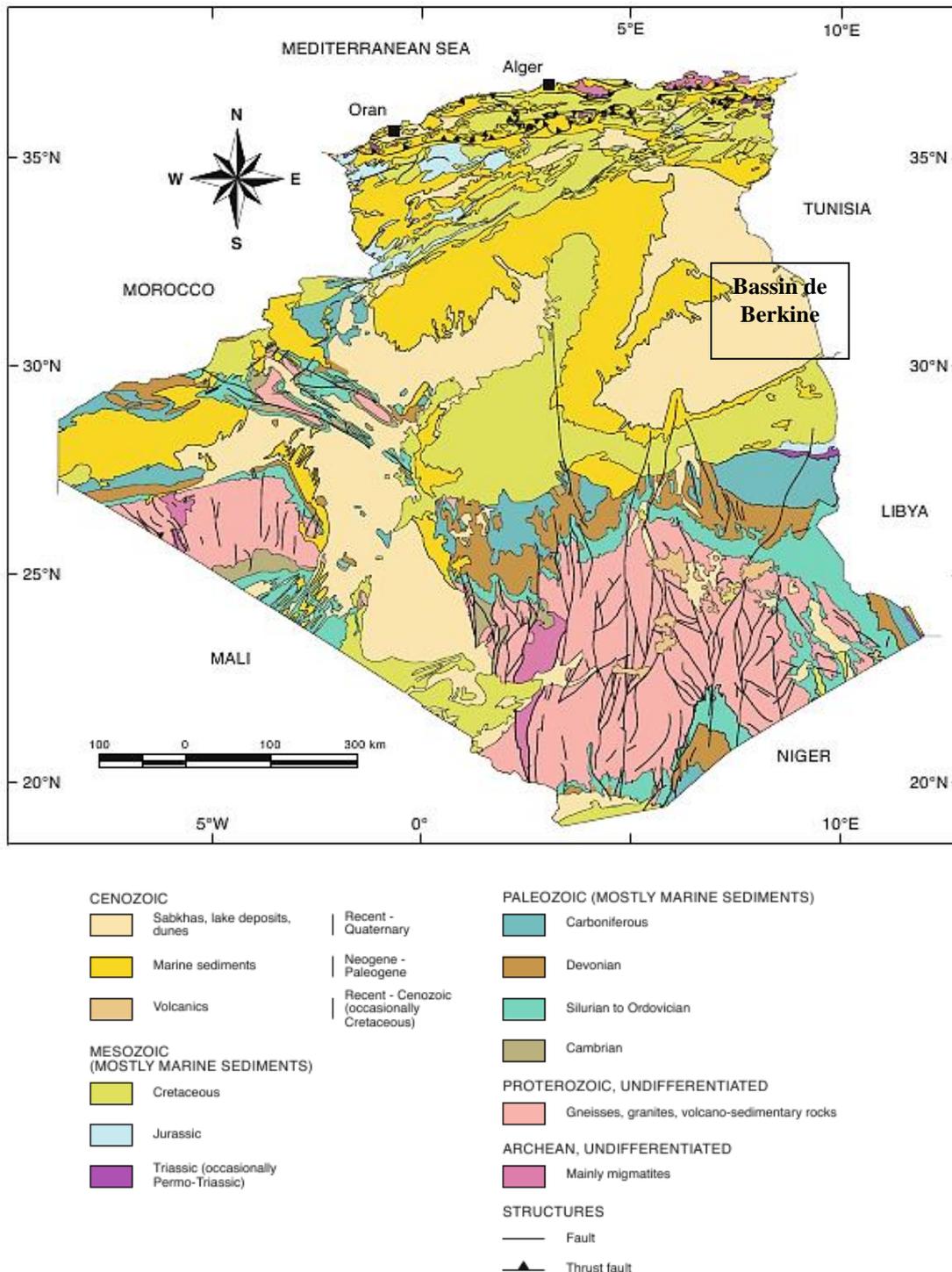


Fig. 22 – La carte géologique de l’Algérie

(<https://gprgindonesia.wordpress.com/2013/07/04/geological-setting-of-algeria/>).

1. 3. Evolution tectonique

Selon les travaux de (Boudjemaa, 1987), le contexte structural globale est les résultats d’une succession de plusieurs phases tectonique depuis le paléozoïque ont été enregistrés dans l’histoire de bassin de Berkine.

1.3.1. Phase Panafricaine

Après la collision entre le craton rigide d'Afrique de l'Ouest et le craton d'Afrique de l'Est le plus plastique, des accidents majeurs subméridiens verticaux se sont produits, qui peuvent être observés sur l'ensemble du socle précambrien, et ils peuvent être interprétés comme des déplacements horizontaux allant jusqu'à 100 kilomètres. Dans ces failles majeures, se conjuguent des réseaux de failles de direction NE-SW et NW-SE, avec une contrainte maximale orientée E-W.

1.3.2. Distension Cambro-Ordovicienne

Il s'agit d'une phase d'érosion sévère, qui a nivelé les reliefs et la structure, l'épaisseur et les changements de faciès des deux côtés de l'accident NS, ce qui peut être visible au niveau de la série Cambro-Ordovicien, indiquant que la partie ouest-africaine plus rigide et la partie orientale la moins rigide, il existe une instabilité tectonique accompagnée de sédimentation (Boudjema, 1987)

1.3.3. Compression tectonique

Il s'agit de l'étape de compression le long de la faille N-S précédemment élargie, formant la structure actuelle du bassin.

1.3.4. Distension Silurienne

C'est le stade où la calotte glaciaire fonte, formant de l'argile noire radioactives.

1.3.5. Compression Calédonienne

Durant cette phase il a un soulèvement général du bassin, et suivi d'une période des dépôts détritiques, survenant entre le Silurien et le Dévonien

1.3.6. Phase Dévonienne

- **Au Dévonien inférieur:** Il y a été caractérisé par des variations dans les épaisseurs et de faciès préexistants le long des axes structuraux subméridiens qui présentent une phase distensive.
- **Au Dévonien moyen et supérieur :** il a été connu par une phase d'érosion appelé discordance Frasnienne.

1.3.7. Discordance Hercynienne

Elle se caractérise par l'arrêt de la sédimentation carbonifère et des dépôts de types lagunaires (argile rouge et gypse). Cette étape tectonique a principalement affecté

la partie nord du bassin de Berkine, où l'érosion carbonifère était presque complète. La discordance hercynienne (DH) a rendu possible le contact des roches mères du Trias, du Silurien et du Frasnien. Les mouvements à ce stade jouent un rôle très important dans la structuration de tous les bassins de la plateforme saharienne.

1.3.8. Phase Autrichienne

Il s'agit d'une phase compressive de direction E-W qui a provoqué une inversion structurelle, au cours de laquelle les failles normales générés dans la séquence triasique liasique ont été inversés.

1.3.9. Phase Pyrénéenne

Il s'agit d'une étape de direction nord-ouest-sud-est, donnant naissance de plusieurs anticlinaux.

1.4. Géodynamique du bassin

Le bassin de Berkine est subdivisé en trois grandes régions pétrolières: la dépression triasique du Sud-Est, la dépression de Dahar et de Berkine (Figure 23).

1.4.1. La dépression Triasique Sud-Est

La dépression est située la cuvette saline de Dorbane et son extension sud-ouest Elle présente la zone entre Emguid Hassi Messaoud jusqu'à l'Ouest et l'Est de bassin du Berkine. Elle est caractérisée par l'évolution des TAGI et TAGS dans les réservoirs du paléozoïque de la plupart des champs pétrolifères découverts (Rhourde Adra, Hassi Chergui, Gassi Touil, Rhourde Nouss, Gassi El Adem, et Rhourde Chouf). Ces champs pétrolifères sont généralement associés aux réservoirs du Trias et au quartzite Hamra. Le pétrole, le gaz naturel et les condensats sont associés à une grande structure reliée par un système de faille inversée à un grand nombre de rejets

1.4.2. La dépression de Dahar

Cette dépression mésozoïque est très vaste, comme structure c'est monoclinale qui plonge vers le nord. Il se poursuit dans le nord jusqu'au bassin Melhill. Il représente la limite nord du bassin de Berkine. De plus, la zone est caractérisée par une érosion hercynienne et taconique. Le développement structurel a une tendance nord-est-sud-ouest. La séquence a été imposée sur l'ancienne séquence du Silurien au Combro-Ordovicien.

1.4.3. La dépression de Berkine

La dépression de Berkine est caractérisée par d'épaisses dunes de sable recouvrant l'erg oriental. Le bassin de Berkine est une très grande dépression datant du Paléozoïque, avec son socle situé au-dessus de 7000m. Sa partie centrale a été affectée par l'érosion hercynienne. Les séries carbonifères terminales sont préservées.

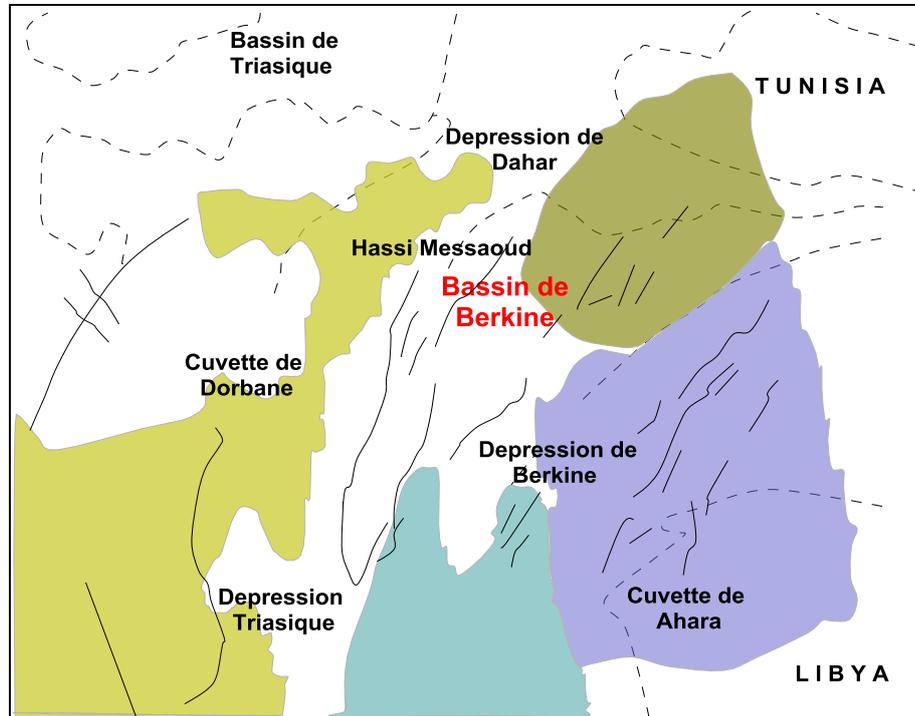


Fig. 23 - Les éléments structuraux du bassin de Berkine (Azoug et al. 2007).

1.5. Stratigraphie

La série stratigraphique est principalement composée par de dépôts mésozoïques qui reposent en discordance sur le Paléozoïque. Et à la fin, un faible épandage de type détritique d'âge tertiaire repose en la discontinuité sur le mésozoïque (Fig. 24)

1.5.1. Le Socle

Pour le socle est composé par des roches cristallophylliennes d'âge précambriennes sur lesquelles reposent en discordance les formations paléozoïques.

1.5.2. Le Paléozoïque

Celui-ci est constitué par :

a) Le Cambrien

Ces dépôts reposent directement sur un socle précambrien, représenté par des grés et des quartzites à passées conglomératiques, avec une épaisseur moyenne de 300m.

b) L'Ordovicien

Il est déposé sur le cambrien est caractérisé par une alternance des grès et l'argile

c) Le Silurien

Il est caractérisé par des argiles noirâtres très riches en matière organique, qui donne une roche mère très fertile, qui est couvert par un importants dépôt argilo-gréseux.

d) Le Dévonien : il est constitué de :

- **Dévonien inférieur:** IL est composé d'une alternance argilo-gréseuse.

- **Dévonien moyen**

Il est caractérisé par un dépôt argileux qui comprend des intercalations de niveaux carbonaté.

- **Dévonien supérieur**

Il est caractérisé par les argiles à passage de marnes de calcaire et de dolomies, au sommet on trouve des grès et des silts.

1.5.2. Le Carbonifère : Il est constitué par:

- **Carbonifère inférieur**

Il est caractérisé par des argiles à passés des silts et de calcaire.

- **Carbonifère supérieur**

Qui a été représenté par des alternances des grès et des argiles sauf on peut observer dans la partie centrale des dépôts carbonatés.

1.5.3. Le Mésozoïque**a) Le Trias**

- **Le Trias Argilo- Gréseux inférieur (TAGI)**

Qui présent l'objectif de notre étude, il est composé par des grès de couleur gris blanc à blanc et gris vert, fin à moyen parfois grossier de forme subarrondi à subanguleux moyennement consolidé avec quelques intercalations des argiles de couleur brun rouge et parfois gris-vert tendre à indurée silteuse, avec la présences des traces de pyrite.

- **Le Trias Carbonaté**

Il est composé par des argiles de verte à gris verdâtre et brun rouge, tendre à indurée, silteuse localement dolomitique avec des passages de dolomie de couleurs blanc beige.

- **Le Trias Argilo- Gréseux supérieur (TAGS)**

Le TAGS, il est constitué par des grès de couleur gris à blanc, très fin,

siliceux à silico-argileux, avec une dureté élevée à intercalations des argiles.

1.5.3.2. Le Jurassique

Il est composé par le Lias, le Dogger lagunaire, le Dogger argileux et le Malm

1.5.4. Le Cénozoïque

a) Le Mio-Pliocène

Il est constitué par des sables de couleur blanc à translucide, rougeâtre, jaunâtre, de granulométrie fin à moyen, parfois grossier et de forme subarrondi à arrondi avec des intercalations des argiles brun à rouge, et avec des bancs de calcaires argileux.

b) Le Quaternaire : il est constitué par des sables représentant des placages dunaires.

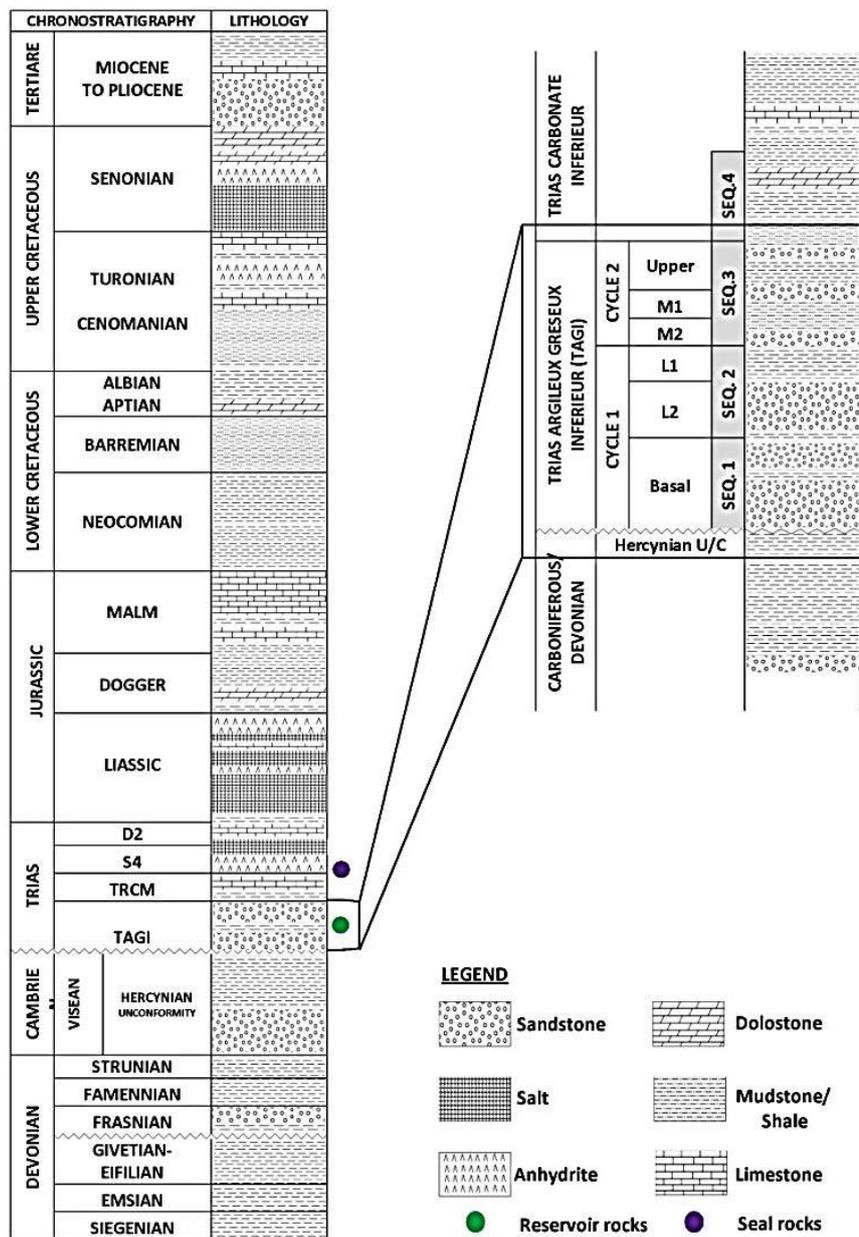


Fig. 24 - Log lithostratigraphique du bassin de Berkine modifié (Rossi et al., 2002).

1. 6. Le système pétrolier du bassin de Berkine

Le système pétrolier du bassin de Berkine a été décrit par de nombreux auteurs, dont (Askri, 1995 ; Echikh, 1998 ; Lassouani, 1996 ; Ghribi, 1999), qui ont été subdivisé le système en quatre principaux éléments: roches mères, roches réservoirs, roches couvertures et les pièges.

1. 6. 1. Roches mères

Les études géochimiques montrent l'existence de deux principales roches mères dans le bassin de Berkine, qui sont les dépôts argileux du Silurien et du Dévonien supérieur (Fig.25).

a. La roche mère silurienne

Le Silurien argileux est presque présent dans tout le bassin, à l'exception dans les bordures Ouest et au Nord où il a été affecté par l'érosion hercynienne (Fig. 25). L'épaisseur varie de 300 à 400 m dans la partie centrale et dans la zone la plus subsidente, il se trouve sur l'axe de Hassi Chergui-Bridès ou son épaisseur peut atteindre les 700 m. Le silurien argileux est subdivisé en deux parties :

- Une partie basale formée par des argiles noires à Graptolites riches en matières organiques et elles sont marquées par un pic radioactif important, avec une épaisseur moyenne varie de 10 à 20 m.
- Une partie supérieure moins radioactive que la partie basale.

La maturation de la roche mère silurienne est située en phase à gaz sur la majeure partie du bassin, particulièrement vers le Sud-Est elle a été marquée par la présence de l'huile.

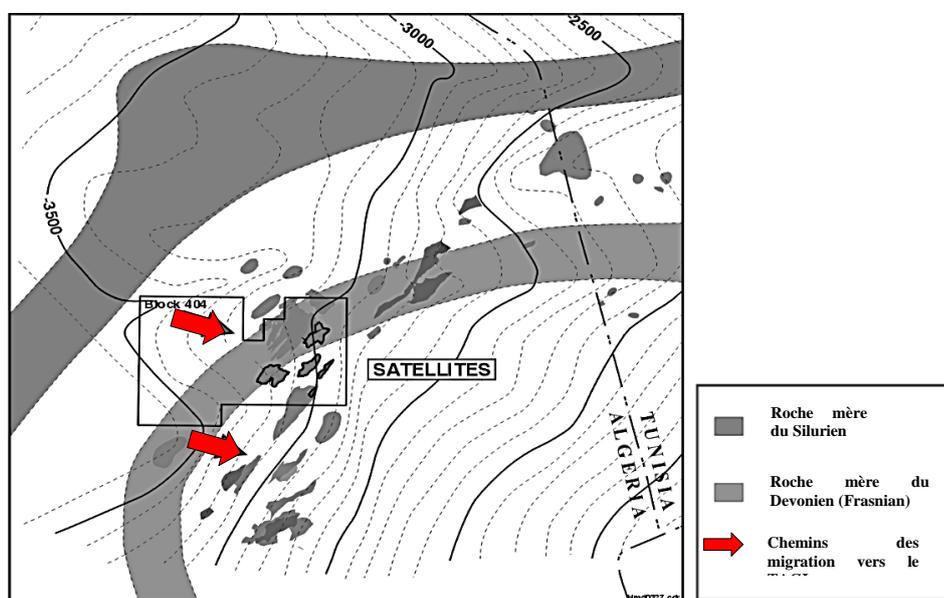


Fig. 25 – Roches sources et migration vers le TAGI (Echikh, 1998).

b. La roche mère dévonienne

Elle a été localisée dans le Dévonien supérieur et subdivisée en deux zones :

- *La zone inférieure* radioactive du Frasnien donne les meilleurs potentiels avec une épaisseur variant de 150 à 300 m.
- *Une zone supérieure* du Frasnien-Famenien d'une épaisseur allant de 50 à 500 m. La maturation à l'actuel est en phase à gaz dans la partie centrale du bassin.

c. La roche mère carbonifère

Les analyses géochimiques ont montré que le niveau d'argiles a un bon potentiel pétrolier; et la matière organique est de type III favorable à la génération du gaz.

d. Les autres roches mères

La roche mère ordovicienne pour cette roche mère la matière organique est localisée dans les argiles micro-conglomératiques de l'Ordovicien supérieur.

1.6.2. Roches réservoirs

Les formations gréseuses du Paléozoïque et du Trias argilo-gréseux constituent les principaux roches réservoirs dans le bassin de Berkine ([Echikh, 1998](#)).

a. Les réservoirs du Paléozoïque

- **Le réservoir cambrien** : il atteint jusqu'au les bordures Nord (môle de Dahar) et Sud-Ouest du bassin avec des faible profondeurs ; son épaisseur varie de 500 à 600 m. Ces réservoir sont formés par les niveaux Ri et Ra où la couverture est assurée par les argiles ordoviciennes d'El Gassi, avec une porosité moyenne variée de 4 à 11 %.
- **Le réservoir ordovicien** : est représenté généralement par les quartzites de Hamra. Surtout dans la partie Sud-Est du bassin qui ont révélé leur importance. Sur l'axe de Gassi Touil- Nezla leur épaisseur maximale est de 325 m, avec une porosité varié de 5 à 10 %.
- **Le réservoir silurien argilo-gréseux** : il est représenté par les unités M, A et B. il correspond à la partie supérieure du Silurien ; elle est subdivisée en cinq unités M1, M2, A, B1 et B2 qui forment la partie inférieure et moyenne du réservoir " F6". Ces réservoirs produits de l'huile et du gaz à condensât à Rhourd Nous et à Rhourde Adra (Fig. 26).
- **Les réservoirs du Dévonien inférieur** : ils sont formés par des grés massifs de la base du Gédinnien ainsi que le Siégenien et l'Emsien ; (alternance de grés et d'argile).

- **Le réservoir Carbonifère** : (Rhourde El Khrouf, (RKF)) : il forme un nouvel objectif pétrolier dans le bassin de Berkine avec des excellentes caractéristiques pétrophysiques, où la porosité moyenne varie entre 8 et 17 %.

b. Les réservoirs du Mésozoïques

Au Mésozoïque, seuls les réservoirs du Trias qui peuvent être considérés comme potentiels pétroliers. Il existe trois principaux réservoirs: le TAGI (Trias argilo-gréseux inférieur qui repose directement sur la surface de la discordance hercynienne, le TAC (Trias argilo-Carbonaté), qui représente la partie médiane) et le TAGS (Trias argilo-gréseux supérieur) constitue la partie supérieure.

- **Le TAGI** : le Trias argilo-gréseux inférieur existe dans tout le bassin à l'exception sur la bordure Sud-Ouest, il présente hiatus. L'épaisseur du TAGI est de 100 m dans la partie central et ouest et varie entre 10 à 200 m avec la présence des roches éruptives dans la partie nord-ouest, qui détruit toutes ses qualités réservoirs.
- **Le TAC** : Le Trias argilo-carbonaté est d'origine lacustre, il est constitué par les argiles de couleur brun rouge dolomitique et silteuse, avec des passages de dolomie microcristalline et de grés fins siliceux à silico-argileux, il est aussi connue par l'existence des roches éruptives dans la région de Nezla.
- **Le TAGS** : Le Trias argilo-gréseux supérieur, il a une extension à la limite de la dépression triasique Sud-Est. Il est de provenance fluviale et formé de grés couleur blancs, silico-argileux un petit peu anhydritique, avec un épaisseur est de 200 m.

1.6. 3. Roches couvertures

La série salifère du Trias-Lias (S4) a donné une bonne couverture régionale pour l'ensemble des réservoirs triasiques connus. Sur le môle d'Ahara, l'équivalent du (S4) se manifeste sous un faciès argileux et peut assurer une couverture étanche comme comme le cas de la région de Hamra et El Ouar Sud, cependant il disparaît au-delà. Elles existent encore des couvertures intermédiaires efficaces comme le Trias carbonaté et le Trias argileux, à l'intérieur du TAGI entre les différents corps de grés, il y a un développement des argiles s'appelle de plaine d'inondation (paléosols), ce qui ont donné les réservoirs multicouches du TAGI, par exemple le cas à Rhourde Er Rouni, Wadi El Teh....

1. 6. 4. Les pièges

Les accumulations des hydrocarbures connus (*Wadi Teh, Bir Berkine, Rhourde El Khrouf, Menzel Ledjmet...*) dans la région sont liées à la présence des pièges structuraux

de types anticlinaux et mixtes, généralement elles sont délimités par des failles normales d'orientation NE-SW. Cependant, le potentiel des pièges est lié aux les déplacements et les rejets de failles qui expliquent le taux de remplissage et les positions des différents plans (huile/eau). (Fig. 26).

Ces failles normales assurent les conduits verticaux et les interconnexions entre les différentes formations des réservoirs ainsi que la superposition des couches entre la roche mère et la roche réservoir par les chemins de migration latérale.

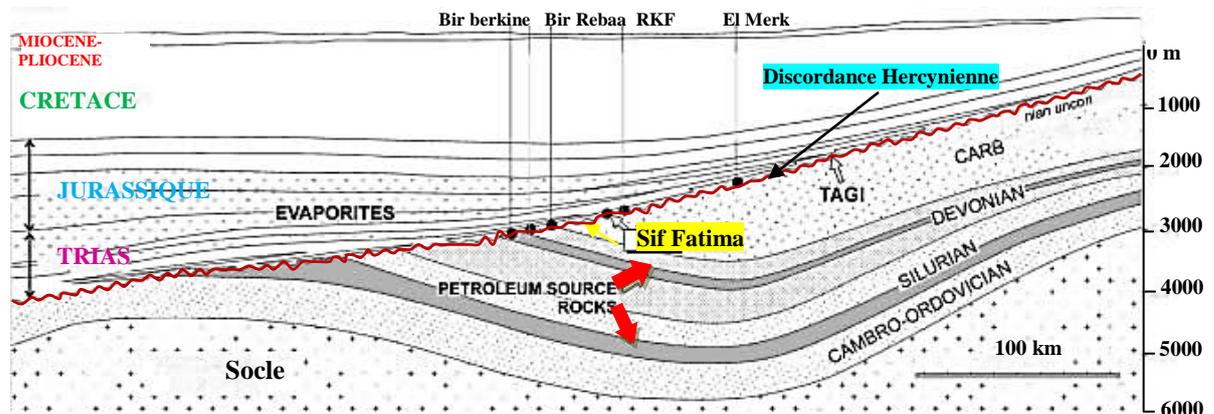


Fig. 26 -Sources et migrations dans le bassin de Berkine (Boote et al., 1998).

2. Géologie Locale du Champ de Sif Fatima

2.1. Situation Géographique

Le champ de Sif Fatima est situé dans le bloc 402 nord-est du bassin de Berkine, à presque de 100 Km au sud-ouest du gisement d'El Borma. Ce bloc couvre environ une superficie de 103 Km² (Fig. 27). Géographiquement, la zone d'étude est localisée entre les latitudes 31°-32° Nord et longitudes 8°-9° Est.

2.2. Synthèse Géologique et Sédimentologique

La région de Sif Fatima est limitée par:

- Bire Rabaa Nord au nord.
- Rhoude Debdaba au sud.
- Bir Berkine à l'ouest.
- Bir Sif Fatima à l'est.

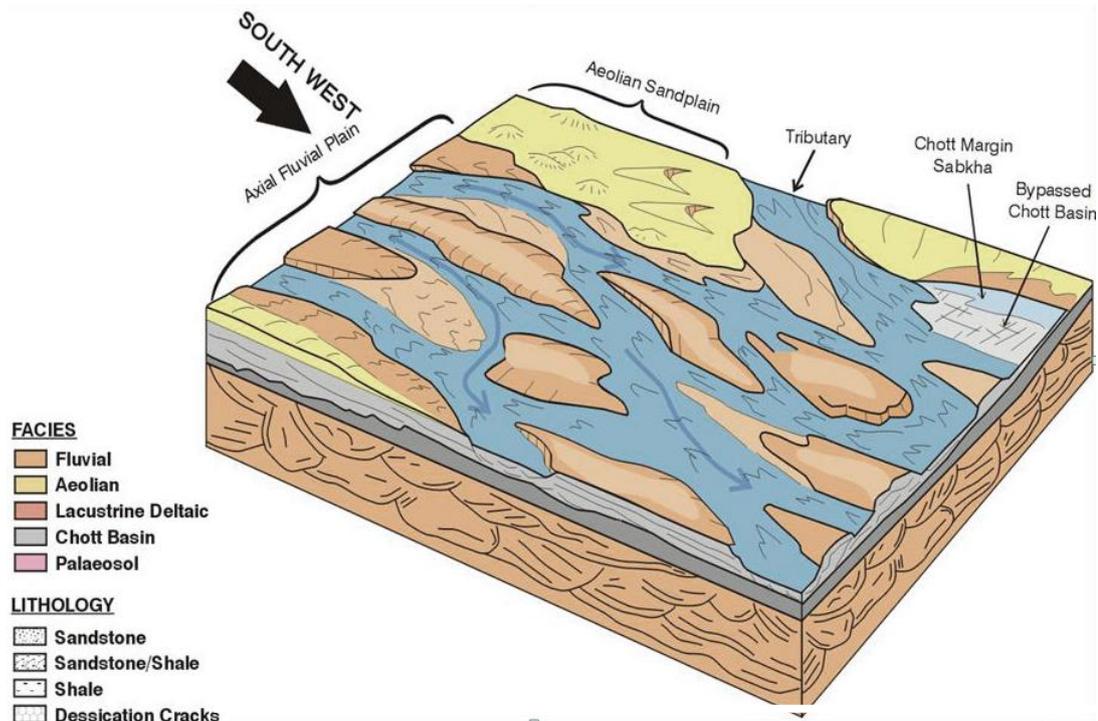


Fig. 28 - Plan et directions des dépôts fluviatiles (Scott & Wheller, 1999).

Le réservoir de TAGI est constitué en trois principales unités (Fig. 28): le TAGI inférieur, TAGI moyen, et le TAGI supérieur, elles même subdivisées en neuf sous couches. Les corrélations entre les puits dans le TAGI sont basées sur l'hypothèse que les deux principaux niveaux d'argile M2 et U2 qui sont régionalement continus dans le champ – alors Ils ont été pris comme un horizon repère.

2. 1. 1. - TAGI inférieur. La partie inférieure TAGI L (Fig. 29), qui se repose directement sur la discordance hercynienne, a été subdivisée en deux niveaux :

- Un niveau d'argile.
- Un niveau de grès, caractérisé par les dépôts fluviatiles, avec un epaisseur tres petit en moyenne de 1.3 mavec des caractéristiques pétrophysiques sont relativement faibles.

2. 1. 2. - TAGI moyen : La couche du TAGI moyen a été subdivisée en quatre sous couches.

■ **M1c, M1b, et M1a :** Ces niveaux sont composés par des grès avec des caractéristiques moyennes. les épaisseurs sont différentes et dépendent selon la distribution locale des chenaux.

2. . 3. - TAGI supérieur (Fig. 29) : La couche du TAGI supérieur est constituée de quatre sous couches.

- **U1a** (sous couche 6), elle est composée par de grès des dépôts fluviatiles avec quelques traversé par les dépôts éoliens. Latéralement, elle est bien prolongée, avec une épaisseur moyenne varie de 2,6 à 10 m.
- **U1b** (sous couche 7), c'est un niveau de grès avec des qualités de réservoir moins bonnes dans le sud-est. Il enferme pareillement quelques dépôts deltaïques intercalés, avec une épaisseur moyenne de 2,9 à 11 m.
- **U2** (sous couche 8), est un niveau d'argile. Il est remarqué dans toute la région, et il est caractérisé par sa petite épaisseur qui variée entre de 0.15 à 3.96 m.
- **U3** (sous couche 9), cette sous couche connue par une épaisseur différent ou les premiers niveaux de grès sont intercalés avec les dépôts deltaïques, avec une épaisseur moyenne de l'ordre de 5,03m.

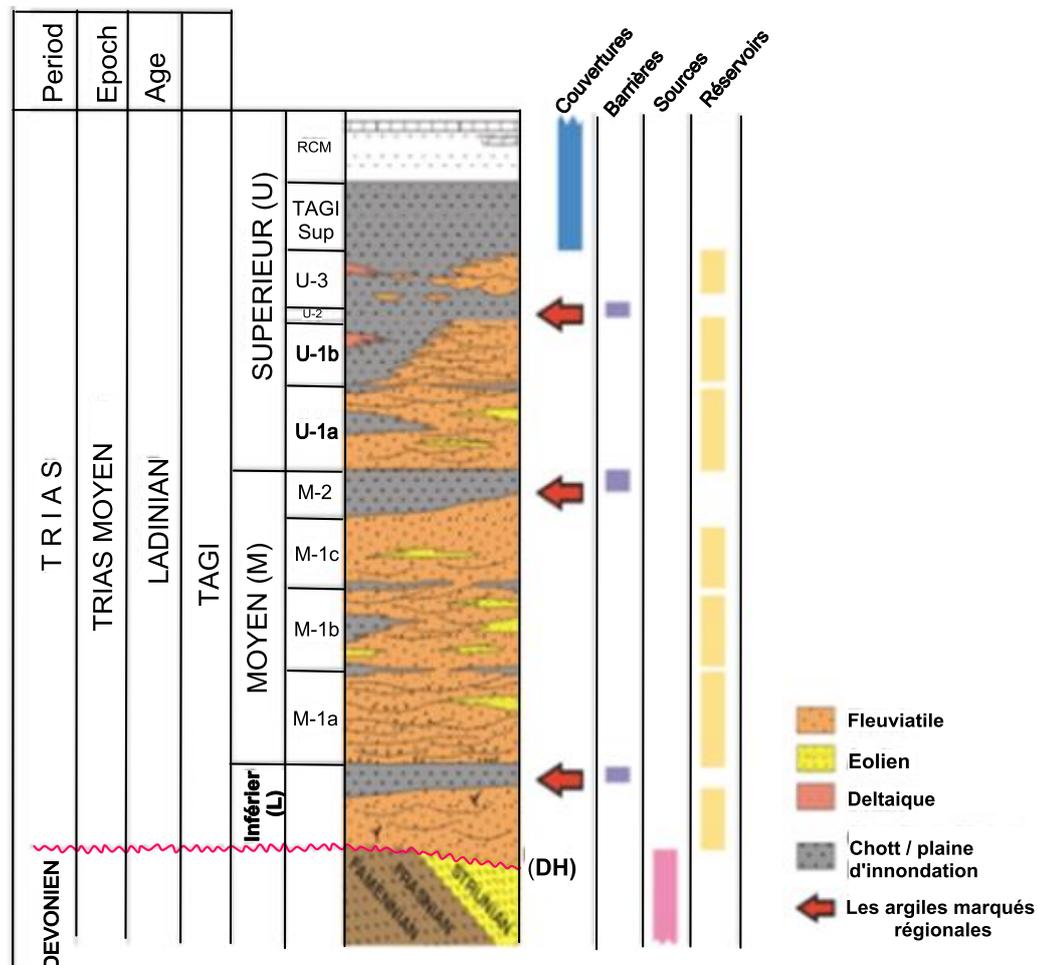


Fig. 29 - Colonne stratigraphique du champ de Sif Fatima (Azoug et al., 2007).

2. 3. Tectonique Locale

La région de Sif Fatima est généralement défini par une tectonique cassante relativement intense, connues par le nom des failles majeurs d'Ourhoud-Sif Fatima, pour les failles secondaires sont pratiquement des mêmes directions (nord-est, sud-ouest) (Fig. 30).

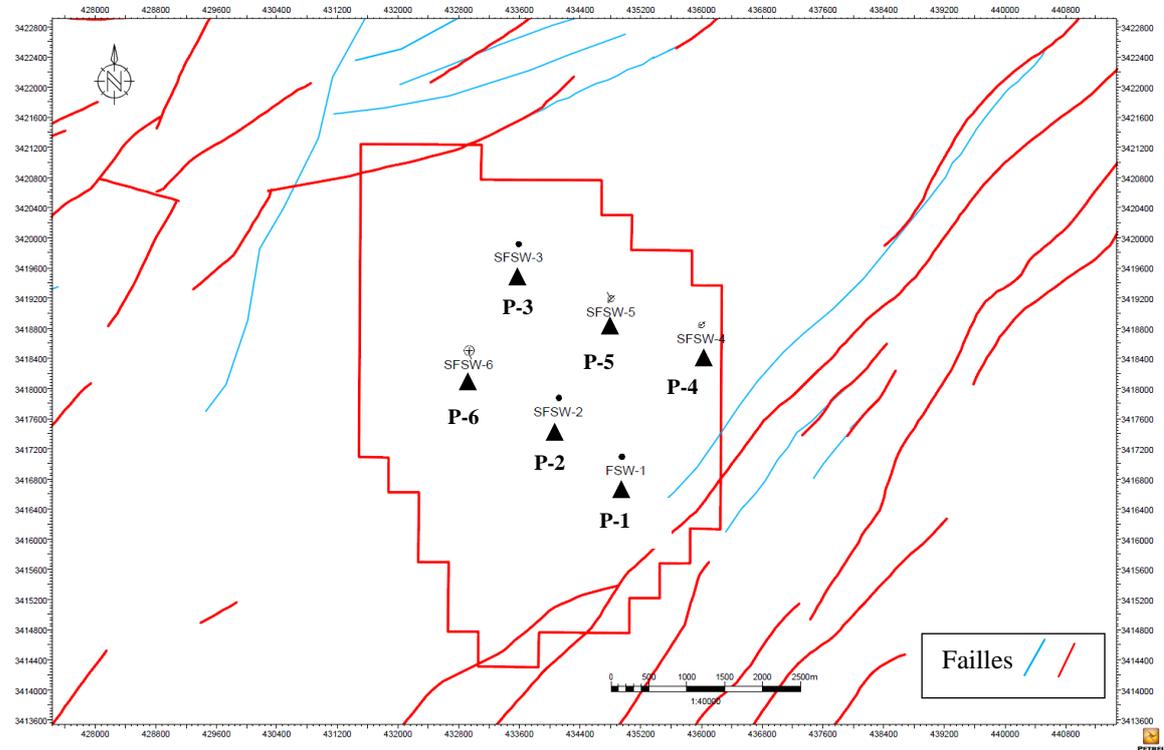


Fig. 30 - Carte Structurale au Toit de TAGI réservoir du secteur Sif Fatima. (Rapport de Sonatrach 2020).

À partir d'analyse des sections sismiques on a distingué: Les marqueurs sismiques représentés par le toit du trias salifère (S4) et le trias carbonaté (TC) et à priori sont les marqueurs les plus continus et énergétiques dans la région. (Fig. 31).

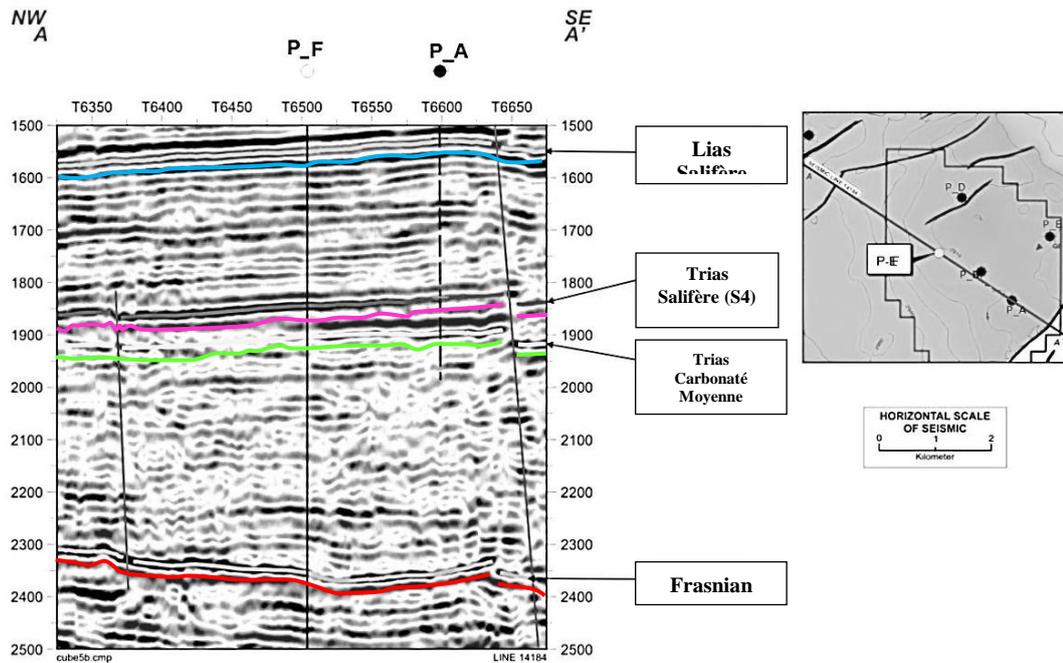


Fig. 31- Coupe sismique du secteur Sif Fatima (Rapport de Sonatrach 2020).

2. 4. Le potentiel pétrolier de la zone d'étude

2. 4.1. Roche mère: Les principales roches mères de la région sont:

- Les argiles du frasnien.
- Les argiles du dévonien qui sont fortement productives .

2.4.2. Roche réservoir: Le Trias Argileux Gréseux Inferieur (TAGI) est le réservoir principal dans la région est qui est subdivisé en trois niveaux qui sont de bas en haut.

a) TAGI basal : composé par:

- Des argiles de couleur brun rouge et gris vert, quelquefois de couleur gris clair indurée légèrement silteuse localement micacée .
- Des grés de couleur blancs, moyennement fin parfois à un couleur blanc verdâtre, friable à moyennement consolidé de forme sub-anguleux à sub-arrondi.
- grés gris-de couleur blanc à gris à clair moyen de granulométrie grossier, de formes sub-arrondies avec des passages d'argile verdâtre de quelques millimètres.

b) TAGI moyen : composé par:

- Grés de couleur beige, fin à moyen bien classé de forme sub-anguleux à sub arrondi, et parfois pris la couleur verdâtre.
- Grés de couleur blanc à beige, de formes sub-anguleuses à sub-arrondi friable, micacé, siliceuse et pyriteux avec des fines passages d'argile verte.

c) **TAGI supérieur**: Compose d'une alternance de grés de couleur blanc à beiges parfois verdâtre, de forme Sub-anguleux à sub-arrondi.

2.3.3. Roche couverture : est assurée par:

- Le Trias carbonaté.
- Les argiles radioactives du Trias argilo gréseux supérieur.

3. Caractéristique Pétrophysique de Réservoir TAGI de la Zone Sif Fatima

3.1. Analyse des carottes

Le carottage est la seule méthode pour récupérer les échantillons de réservoir afin de faire des mesures directes et connaître les propriétés des roches et des fluides contenus. L'analyse des carottes récupérées peut fournir les données qui permettront la bonne interprétation des réservoirs ou de l'intervalle de formation carottée. elle est divisée en deux catégories; une analyse dite conventionnelle et une analyse spéciale. L'analyse conventionnelle comprend les tests directs qui déterminent les propriétés pétrophysiques suivantes:

- Une caractérisation lithologique et faciologique de la roche.
- Mesure de la porosité de la roche qui déterminera le volume réel des fluides dans le réservoir.
- Mesure de la perméabilité de la roche pour estimer la capacité de transmettre les fluides et le pouvoir de production du réservoir.
- Mesure de la saturation de la roche par les fluides est indiquée la présence et le type d'hydrocarbure dans la roche.

Les autres informations qui ont été obtenues par l'analyse conventionnelle comprennent le type de fluides produits et le contact entre eux.

L'analyse plus poussée des carottes par des tests plus sophistiqués que ceux réalisés dans l'analyse classiques.

Le (TAGI) c'est l'objectif principal de la structure de Sif Fatima caractérisé par une profondeur moyenne **2934 m**, d'épaisseur moyenne **98 m**. et de bonne porosité **18-20 %**. Et perméabilité **800-1000md**. Les découvertes d'huiles dans les réservoirs de TAGI ont été mises en évidence un gisement multicouche associé à des Grès fluviatiles. On a obtenu la description générale des carottes ; après l'observation et l'analyse de toutes les carottes (Fig. 32):

Des alternances de grés de couleur gris, gris-beige à gris-verdâtre et parfois blanc à blanc-beige; fin à très fin à rarement moyen; moyennement bien classé, et de

forme sub-arrondi et sub-anguleux; avec des argiles brun-rouge, gris-verdâtre et parfois sombre plus ou moins consolidé indurée Micacée, Silteuse, Carbonaté avec présence localement de Pyrite, de Chlorite et rarement de Quartzite avec passage cm à dm de siltstones. Présence des quelques fissures horizontale et oblique à stratification inobservable avec présence très peut fortement des fissures verticale. Ciment argileuse et parfois peut fortement de argilo-bitumineux, siliceux à silico-bitumineux.

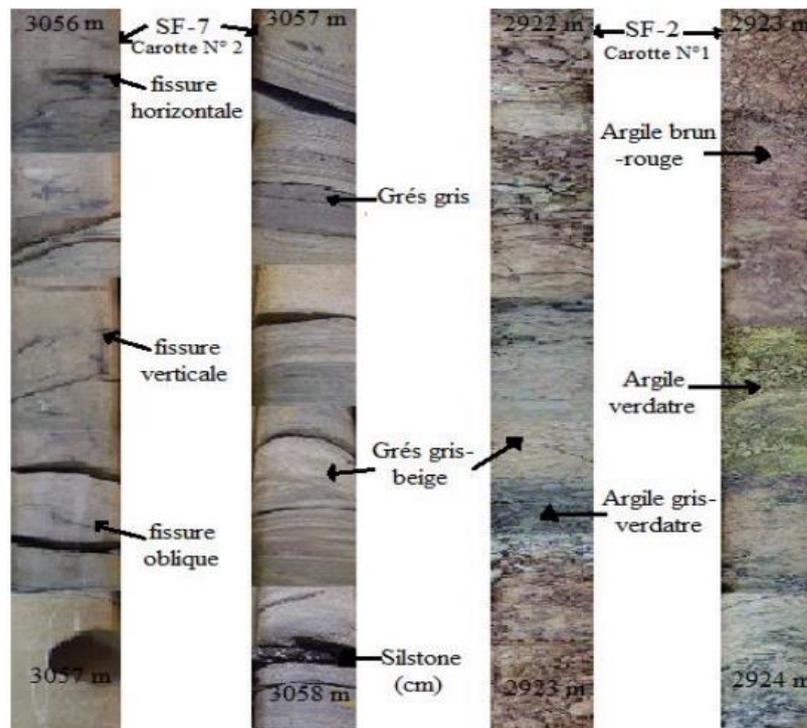


Fig. 32- Photos de quelques carottes à différents intervalles du réservoir TAGI.

3.2. Synthèse pétrophysique du réservoir TAGI

L'exploitation optimale d'un gisement des hydrocarbures exige de connaître aussitôt que possible:

- Les propriétés du réservoir (porosité, perméabilité).
- La quantité des hydrocarbures en place.

Le réservoir du secteur Sif Fatima (bloc 402b) ayant une excellent propriétés (porosité et perméabilité) et quantité des hydrocarbures capable pour d'exploitation optimale ultérieurement. À partir des données disponibles (interprétation diagraphie et résultats d'analyse des carottes); Ces caractéristiques en générale sont:

a. Puits P_A:

Le top de réservoir (TAGI) 2962.5m L'épaisseur 86m, en caractérise par:

TAGI supérieur-moyen: la porosité moyenne est de 14 %, TAGI inférieur: la porosité moyenne est de 17.95 %.

b. Puits P_B:

Le top de réservoir (TAGI) 2966m d'épaisseur 77m, en caractérise par:

TAGI supérieur-moyen : L'épaisseur est de 9m ont donné 16.73 m³/h d'huile, La porosité moyenne est de 17 %, la saturation en eau moyenne est de 17.5 %.

TAGI inférieur: L'épaisseur est de 3m ont donné 16.73 m³/h d'huile, La porosité moyenne est de 16 %, la saturation en eau moyenne est de 28%.

c. Puits P_C

Le top de réservoir (TAGI) 2925m d'épaisseur 85m. en caractérise par:

TAGI inférieur: est saturé en hydrocarbures sur une hauteur utile de 23m ont donné 20.17 m³/h d'huile avec une porosité moyenne de 16% et une saturation en eau moyenne de l'ordre de 20%, TAGI supérieur-moyen: la porosité moyenne est de 18%, et une saturation en eau moyenne est de l'ordre de 20%, sur une hauteur utile de 40m ont donné 16.63 m³/h d'huile.

d. Puits P_D:

Le top de réservoir (TAGI) 2925m d'épaisseur 100m, en caractérise par:

TAGI supérieur: la porosité moyenne est de 16%, la saturation moyenne en eau est de 10% sur une hauteur utile de 10m. TAGI moyen: la porosité moyenne est de 18%, la saturation moyenne en eau est de 10% sur une hauteur utile de 15m. TAGI inférieur: la porosité moyenne est de 19%, la saturation moyenne en eau est de 35% sur une hauteur utile de 18m.

e. Puits P_E:

Le top de réservoir (TAGI) 2904m L'épaisseur 109m, en caractérise généralement par: La porosité moyenne 20 %, de perméabilité moyenne 473 md, et de saturation (Sw 22%, So48%, Sg 28%).

3.3. Historique des travaux réalisés sur le bassin de Berkine

Une large activité de l'exploration de ce bassin est continu par la prospection sismique date depuis 1986.

* 43 097 km en 2D sismique acquis entre 1986 et 2004.

* 8 119 km² de sismique 3D acquis entre 1995 et 2004.

L'exploration du bassin de Berkine a été faite par puits. Un total de 215 puits a été foré de 1985 à 2004. (Azoug et al., 2007). Ce bassin a fait l'objet de plusieurs études à cause de la richesse en hydrocarbures. (Turner et al. 2001 ; Rossi et al. 2002 ; McKenna & Hedley 2002; Chelbeb et al. 2000 ; Dieckmann et al. 2000).

- En 1989 Anadarko a signé, avec Sonatrach, le premier contrat d'exploration dans le bassin de Berkine. Le premier puits a été foré en 1991.
- Entre 1993 et 1998, d'autres découvertes sont faites, et c'est le 4 Mai 1998 que le Groupement Berkine-Sonatrach était né (association Sonatrach, 51%- Anadarko, 49%).

Conclusion

La région de Sif Fatima est située dans le grand 'erg oriental du Sahara algérienne, Cette région est couvrée d'une superficie plus de 103Km². Il est délimitée par Bir Rabaa au nord; Rhoudé Debdaba au sud; à l'ouest par Bir Berkine; et Bir Sif Fatima à l'est. Les formations sédimentaires ont été représentées par des dépôts d'âge mésozoïque, reposant en discordance (DH) sur les terrains du paléozoïque, qui sont traversées relativement, par un épandage détritique qui est daté d'âge tertiaire et surmonté en discontinuité avec le crétacé. L'ensemble de la série sédimentaire est couvert par des sables dunaires du miopliocène et du quaternaire.

La région de Sif Fatima est généralement défini par une tectonique cassante relativement intense, connues par le non des failles majeures Ourhoud-Sif Fatima et les failles secondaires sont presque des mêmes directions (nord-est, sud-ouest)

Le réservoir Trias Argileux Gréseux Inferieur (TAGI) c'est le principal objectif de de la structure de Sif Fatima, qui est caractérisée par une séquence verticale d'origine fluviatile, multicouche associé à des grès fluviatiles ont des caractéristiques pétrophysiques appropriées.

Le champ de Sif Fatima est connu par des problèmes de la récupération des carottes à cause des problèmes géologiques (hétérogénéité du réservoir TAGI) et techniques rencontrés durant l'opération de carottage, ce qui résulte des lacunes des informations faciologique et pétrophysique (porosité et perméabilité).

Chapitre III

**Prédiction des facies non
Carottés**

Chapitre III. Prédiction des faciès non carottés

Introduction

Le meilleur moyen pour caractériser et évaluer les réservoirs pétroliers sont les échantillons de carottes ; cependant la récupération des carottes dans les forages pétroliers n'est pas toujours intégrale, à cause des problèmes géologiques et techniques rencontrés durant l'opération du carottage, et sa limitation considérant au nombre de puits dans le champ (Dubois et al, 2007). D'autre part, pendant la modélisation du réservoir, l'identification des lithofaciès dans tous les puits étudiés avère être une phase importante dans les modèles statiques et dynamiques. Par conséquent, des lithofaciès doivent être appropriés dans tous les intervalles non carottés, alors le défi est consisté à trouver une méthode fiable pour estimer les faciès lithologiques dans les intervalles non carottés.

Les réseaux neurones artificiels (ANN) ont été appliqués dans un nombreux domaines pour résoudre une variété de problèmes tels que la classification, l'extraction de dispositif, le diagnostic, l'approximation de fonction et l'optimisation (e.g., Lawrence, 1994).

Cette méthode est étendue aux grandes applications de modélisation de différents paramètres dans les réservoirs pétroliers. Il s'intéresse à la quasi-totalité de la chaîne d'exploitation de l'industrie pétrolière ainsi qu'aux différentes applications sur les données sismiques (Diersen, et al., 2011), la géochimie (Sun et al., 2009), le forage (Kannan et al., 2014), le well testing (Adibifard et al., 2014) et les diagraphies (Salehi et al., 2017). Divers méthodes statistiques comme fonction de discrimination; La technique d'analyse factorielle, le composant principal, le classificateur bayésien naïf, support vector machine, les algorithmes de partitionnement et cluster analysis ont récemment été discutés et appliqués aux données de diagraphie par câble pour la reconnaissance du faciès lithologique (e.g., Szabó, 2011; Szabó & Dobróka, 2017; Sebtosheikh et al., 2015; Asfahani, 2014; Al-Mudhafar, 2015). Au cours des dernières décennies, de nouvelles méthodes basées sur l'ANN ont été développées et appliquées avec succès sur les réservoirs de pétrole pour résoudre les problèmes de lithofaciès non-carottés (e.g., Jianhua et al., 2019; Aliouane et al., 2012).

Dans ce chapitre, le réseau neuronal perceptron multicouche (MLPNN) et cluster analysis sont les techniques choisies pour prédire les lithofaciès non carottés dans les puits étudiés. Les deux méthodes seront comparées pour leurs puissances et leurs performances.

Cluster analysis qui est une technique statistique de calcul, consiste à diviser la base de données en différents paquets homogènes (Kiselev and Baikalskaya, 2011), en ce sens que les données de chaque subdivision ont des caractéristiques communes et de faire l'écart du logarithme par faciès (Ameur-Zaimeche et al., 2014a), l'attribution des faciès prévus proposés en obtenant la valeur minimale des erreurs.

Par ailleurs, cette étude vise à générer un faciès lithologique synthétique sur 333 m qui n'ont pas été carottés du réservoir fluviatile de Trias Argileux Gréseux Inférieur (TAGI). Nous notons que seules quelques études ont été menées pour étudier les capacités et l'utilité du réseau neuronal artificiel dans l'estimation des électrofaciès en utilisant uniquement des variables de diagraphie par filaire nucléaire dans des réservoirs détritiques hétérogènes, ainsi, la présente étude on cherche à développer nos compréhensions dans ce sujet et se concentre sur l'application du modèle MLPNN pour prédire les lithofaciès en utilisant quatre variables de diagraphie nucléaires comme prédicteur: rayons gamma, densité, thorium et potassium combinés avec une description codée des lithofaciès existants, représentant les données d'entrées, pour la comparaison avec les résultats de MLPNN et cluster analysis a été réalisé et les performances des modèles seront vérifiées à l'aide de l'erreur quadratique moyenne (RMSE), de l'erreur absolue moyenne (MAE) et du coefficient de corrélation (R).

1. Préparation de la Base des Données

Cette étude comprend cinq puits dans le réservoir TAGI du champ pétrolifère Sif Fatima, pour lesquels trois puits (W-A, W-B, W-C) avaient été carottées correspondant à 1242 observations lithofaciès à des intervalles de carottage totale de 189 m, et deux autres puits non carottés (W-D, W-E) représentent les lithofaciès manquants avec un total de 333 m correspondant à 2184 observations manquantes, ces intervalles manquants font l'objet de cette étude.

La préparation de la base de données a été basée sur le codage des observations réelles sur les carottes, on donne des chiffres selon les faciès disponibles en respectant le même pas des mesures qui est de 0,1524 m (1/2 pied) des diagraphies calés avec les intervalles carottés. Les lithofaciès ont été subdivisés en quatre (classes) et ont été codés comme suit: (argile = 0), (siltstone = 1), (grès = 2), (grès-argileux = 3). Dans cette étude, quatre types de données de diagraphies des puits ont été impliqués (gamma ray naturelle, densité, thorium et potassium). Le tableau 2 résume les statistiques descriptives de l'ensemble de données, où X_{mean} , X_{max} , X_{min} , S_x et C_v indiquent respectivement la moyenne, le maximum, le minimum, l'écart-type et la variance.

Tableau. 2: Statistique des paramètres de diagraphies (GR, RHOB, THOR et POTA)

Statistique	GR(A)	RHOB(B)	THOR(C)	POTA(D)
Moyenne	81.4866	2.5398	6.9686	2.3319
Erreur-type	1.5608	0.0046	0.1951	0.0440
Médiane	71.7729	2.5395	5.2769	2.2797
Écart-type	31.6040	0.0938	3.9513	0.8917
Variance	998.8144	0.0088	15.6126	0.7951
Minimum	13.1673	2.3604	0.7851	0.0714
Maximum	140.8081	2.7203	15.3682	5.1348
Somme	33409.5027	1041.3347	2857.1118	956.0851

La base de données est divisée en trois sous-ensembles: apprentissage, validation et test, parmi les 1242 données, 870 paires entrées-sorties (70%) ont été sélectionnées au hasard parmi les séquences de données utilisées dans l'ensemble d'apprentissage.

L'ensemble de data pour la phase de validation comprend 186 paires d'entrées-sorties représentant 15% de l'ensemble des données, tandis que les données restantes, qui

consistent en 186 paires de sorties (15%) de l'ensemble de données disponibles, ont été utilisées pour tester les modèles.

L'ensemble de validation est utilisé pour s'assurer qu'il n'y a pas de sur-apprentissage dans le résultat final. Les trois sous-ensembles ont été tirés au hasard des données d'origine avec la fonction *dividevec* dans Matlab, L'échantillon de validation a été utilisé pour garantir la robustesse des paramètres du réseau, pour décider quand la formation doit être arrêtée et pour surveiller les erreurs. Par conséquent, la formation est arrêtée lorsque l'erreur continue d'augmenter. Le sous-ensemble de test a été utilisé pour tester le modèle avec des données entièrement inconnues.

2. Elaboration du Modèle des Réseaux de Neurones Artificiels

La cette méthode est basée sur la technique des réseaux de neurones artificiels, c'est une technique inspirée de la fonction des neurones biologiques. Ils constituent une famille de fonctions non linéaires paramétrées, utilisées dans de plusieurs disciplines (physique, chimie, biologie, finance, géologie, etc.), notamment pour la modélisation de processus complexe. Dans le cadre de notre travail un type particulier a été appliqué: Le perceptron multicouche (MLPNN) (*voir chapitre I*).

Le modèle employé pour résoudre le problème est de perception multicouches avec une algorithme s'appelle (*Retropropagation*) à apprentissage supervisé, le modèle est composé d'une couche d'entrée composé à 4 neurones qui sont les variables explicatifs (*diagraphies nucléaires*), et d'une couche de sortie composée de quatre neurones qui sont les variable à expliquer (les faciès) , et une seule couche cachée. Pour les fonctions de transfer ; de la couche d'entrée vers la couche cachée c'est une fonction sigmoïdale (*Tansig*), et pour la couche cachée vers la couche de sortie et une fonction linéaire (*Purelin*), et l'algorithme d'activation appliqué est de (*Levenberg-Marquardt*). A chaque fois on change le nombre de neurones dans la couche cachée, le modèle à une capacité d'amélioration et d'adaptation.

2.1. L'architecture du réseau de neurones

La couche d'entrée est constituée de quatre variables qui sont les diagraphies les variables explicatifs (*Gamma-Ray, Thorium, Potassium, et la Densité*).et une seule couche cachée, constituée de 4 neurones. La couche de sortie est constituée de quatre neurones qui sont les variables à expliquées, qui sont le Faciès non carotté (argile, silt, grès, et grès argileux) qu'on le cherche dans les intervalles non carottées.

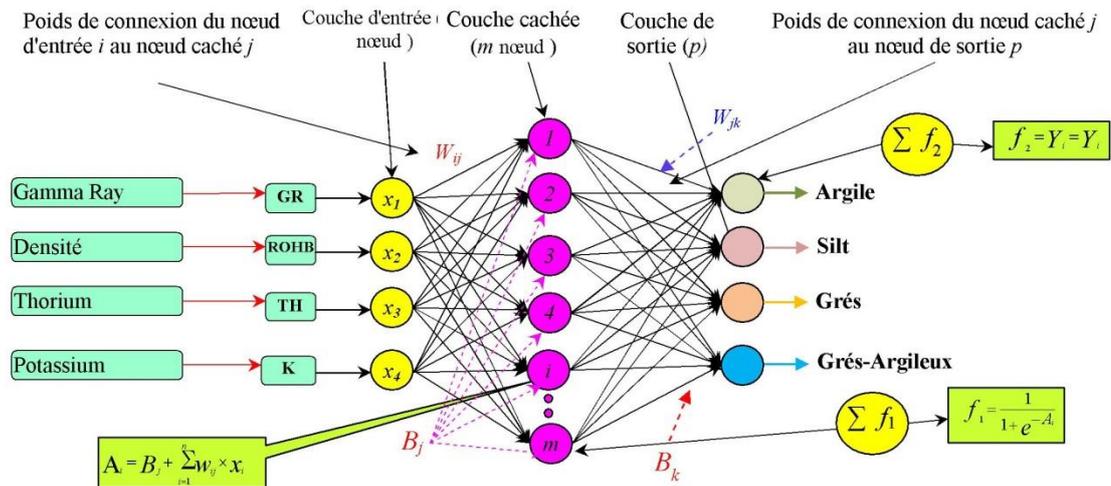


Fig. 33 - Architecture du réseau de neurones élaboré pour reconstituer les faciès (Ameur-Zaimeche et al, 2020).

2.2. Amélioration des performances du réseau

Le modèle MLPNN est composé de trois couches, avec une seule couche cachée en utilisant une fonction d'activation sigmoïde qui montre une meilleure structure de robustesse comme indiqué dans la littérature (Namig & Vugar 2017; Heddami, 2016). La détermination du nombre de neurones cachés est effectuée sur la base d'une règle d'essai et d'erreur (Ramadevi et al., 2012; Sheela & Deepa 2013). Le résultat qui affiche une erreur minimale permet de définir le nombre de neurones cachés. En d'autres termes, la structure de réseau idéale a été recherchée en modifiant le nombre de neurones dans la couche cachée jusqu'à trouver les topologies MLPNN optimales comme indiqué dans le tableau 3.

2.3. Critères de validation

Une fois le modèle construit se pose alors la question s'il rend fidèlement compte de l'échantillon. Pour cela, la modélisation nécessite des critères de validation sur la base desquels, le concepteur juge les performances finales du modèle qui sera retenu. Certains sont numériques et d'autres graphiques (Willmott CJ (1981,1982).in I. Ladlani et al. (2012)).

2.3.1. Le coefficient de détermination (R^2)

Celui-ci évalue le degré d'association entre deux variables et juge de la qualité de l'ajustement des valeurs observées et calculées :

$$R^2 = \left[\frac{\frac{1}{N} \sum_{i=1}^N (Y_{iobs} - \bar{Y}_{obs})(Y_{ical} - \bar{Y}_{cal})}{\sigma_{obs} \cdot \sigma_{cal}} \right]^2 \quad (7)$$

Avec : Y_{iobs} : Valeur observée. Y_{ical} : Valeur calculée par le modèle.

\bar{Y}_{obs} : Moyenne de la variable observée.

\bar{Y}_{cal} : Moyenne de la variable calculée. N : est le nombre de points.

Un coefficient de détermination proche de l'unité signifie que le modèle élaboré est très efficace.

2.3.2. Erreur Quadratique Moyenne, Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^n (Y_{iobs} - Y_{isim})^2} \quad (8)$$

Quand le RMSE est proche de zéro signifie que le modèle est très bon.

2.3.3. Erreur Absolue Moyenne, Mean Absolute Error (MAE)

$$MAE = \frac{1}{N} \sum_{i=1}^N |Y_{iobs} - Y_{isim}| \quad (9)$$

Lorsque le MAE est proche de zéro signifie que le modèle est très bon.

Remarque : On a changé le nombre de neurone dans la couche cachée de 1 à 20 neurones et on a uniquement mis dans quelques changements.

Tableau. 3 : Performances des modèles de MLPNN élaborés.

Nbr de neurones	Apprentissage			Validation			Test		
	R ²	RMSE	MAE	R ²	RMSE	MAE	R ²	RMSE	MAE
1	0.84	0.52	0.28	0.76	0.57	0.30	0.86	0.50	0.27
2	0.46	0.87	0.75	0.07	0.91	0.85	0.12	0.97	0.91
3	0.84	0.52	0.30	0.76	0.58	0.35	0.86	0.51	0.30
4	0.87	0.48	0.27	0.78	0.55	0.30	0.92	0.39	0.23
5	0.84	0.54	0.37	0.74	0.60	0.37	0.85	0.55	0.40
6	0.83	0.55	0.31	0.74	0.59	0.34	0.84	0.55	0.33
7	0.84	0.53	0.32	0.76	0.58	0.35	0.86	0.51	0.31
8	0.85	0.51	0.30	0.76	0.58	0.33	0.88	0.47	0.29
9	0.85	0.51	0.32	0.74	0.59	0.38	0.86	0.50	0.33
10	0.86	0.50	0.28	0.76	0.58	0.33	0.88	0.47	0.27
11	0.86	0.50	0.30	0.76	0.58	0.36	0.88	0.47	0.30
12	0.87	0.49	0.27	0.76	0.58	0.34	0.87	0.48	0.29
13	0.84	0.53	0.37	0.74	0.59	0.36	0.85	0.55	0.40

Les performances du modèle se changent on fonction de l'ajout d'un neurone dans la couche cachée, à de quatre neurones donnent des meilleurs résultats. Les indices numériques du modèle sont : un coefficient de détermination (R²) de 0,87 pour l'apprentissage, 0,78 pour la validation et 0,92 pour le test.

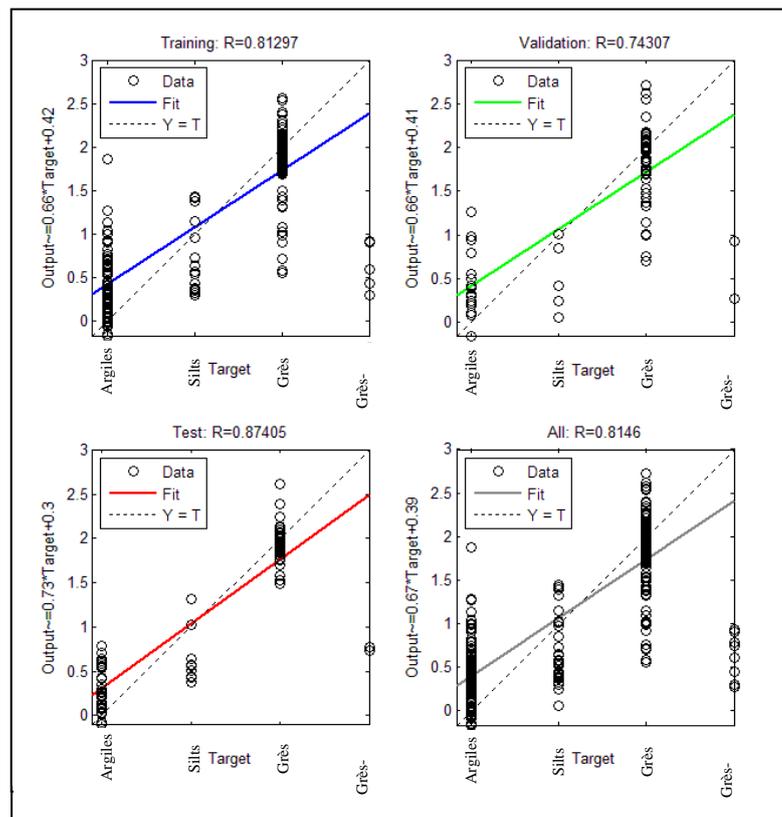


Fig. 34- Régression entre les faciès réels et les électrofaciès estimés dans les étapes de l'apprentissage, de validation et le test.

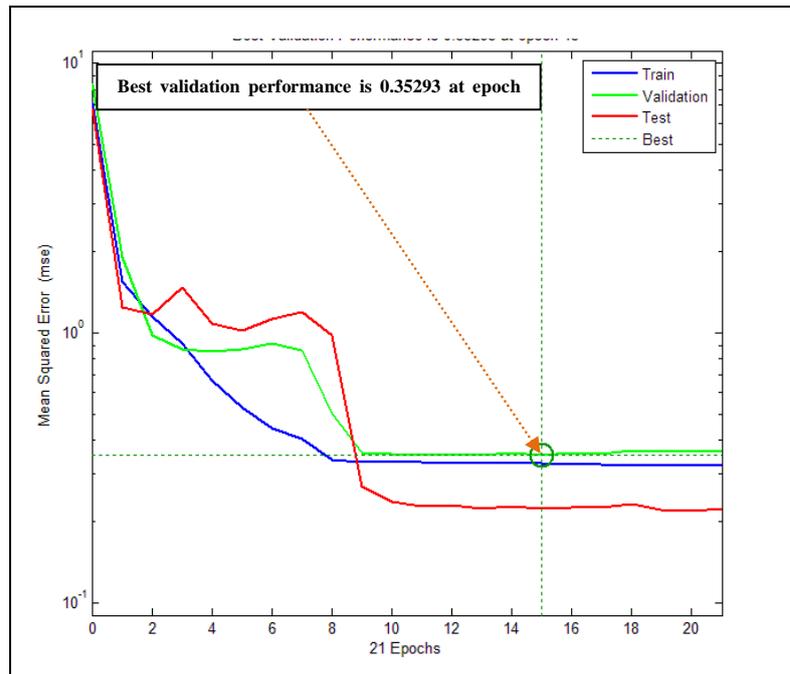


Fig. 35 - Fonction coût de performance des étapes d'apprentissage (ligne bleue), de validation (ligne verte), et du test (ligne rouge).

La figure (35) montre la fonction coût (l'arrêt précoce), qui mesure la distance entre la sortie du modèle et la sortie target, elle est une fonction scalaire qui dépend du vecteur de paramètres du modèle, et les variables de l'ensemble d'apprentissage ; dans le cas des réseaux de neurones, le vecteur de paramètres est constitué par les poids du réseau. Plus la valeur de la fonction de cout est petite, plus le modèle reproduit fidèlement les observations utilisées pour l'apprentissage. Les différents algorithmes d'apprentissage cherchent donc à trouver le point, dans l'espace des paramètres, pour lequel la fonction de cout est minimale et au de là le modèle se diverge et se produit le sur-apprentissage (Ameur-Zaimeche, 2014b).

Les paramètres retenus par le modèle élaboré sont présentés dans le tableau suivant :

Tableau. 4 : Les paramètres retenus des modèles de MLPNN.

N° Neurone	Poids de la couche d'entrée				Neurone de sortie		
	Gamma Ray	Densité	Thorium	Potassium	biais	poids	biais
1	2.95	-36.08	-3.82	25.26	4.60	0.17	
2	-5.50	-47.03	3.11	-0.05	35.25	0.69	
3	-9.86	87.74	2.45	-0.93	-20.27	-0.49	
4	0.59	23.84	-12.31	-13.89	6.41	1.83	
5	8.46	21.89	-1.99	-3.76	-18.02	-0.89	
6	0.19	147.48	0.52	0.37	-53.98	0.34	
7	7.69	-50.16	0.69	12.15	11.37	-1.08	1.99
8	-5.37	93.57	-17.78	-0.82	-18.28	0.39	
9	4.96	60.07	-3.39	-1.54	-32.76	-2.40	
10	15.35	-38.26	15.84	-1.80	25.22	-3.79	
11	2.95	-36.08	-3.82	25.26	0.98	3.06	
12	-5.50	-47.03	3.11	-0.05	62.93	0.41	
13	-9.86	87.74	2.45	-0.93	44.39	-0.71	

3. Application de la Méthode de Classification Cluster Analysis

La technique dite cluster analysis est un groupe de méthodes mathématiques permettant de reconnaître des groupes naturels et significatifs au sein d'un ensemble d'échantillons. Cette méthode qui appartient à la classification non supervisée, vise à diviser les données en groupes ayant des caractéristiques similaires sans aucune connaissance préalable des données étudiées. Les approches utilisées pour construire une structure de groupe de données sont notamment le partitionnement et les méthodes hiérarchiques (Kaufman & Rousseeuw, 1990). En d'autres termes, l'analyse de regroupement implique le développement de la division de certains objets échantillons en subdivisions appelées cluster de telle sorte que chaque paquet soit constitué d'objets similaires, tandis que les objets de cluster différent seraient considérablement différents. La méthode de créer des clusters se rapporte au traitement des données statistiques (Kiselev, 2011) est suffisamment efficace, car étant une méthode analytique, elle n'a pas de jugement subjectif associé à l'analyse visuelle des objets graphiques (Parsaye, 1998; Pospelov, 1988). Il présente de nombreux avantages qui facilitent de nombreuses tâches dans différents problèmes. Dans cette étude, nous avons utilisé une approche en quatre étapes pour prédire aucun faciès de carottage dans le réservoir TAGI: 1. Dans un premier temps, c'est la détermination de la différence entre la valeur moyenne de tous les logs (GR, RHOB, Th, K) et les valeurs actuelles de chaque faciès de codage, ce fait est obtenu l'écart des logs $D(X_i)$ est représenté par Eq. (10).

$$D(X_i) = X_{(i)} - \bar{X}_{(i)} \quad (10)$$

Où $X_{(i)}$ est la valeur des données de diagraphies, $\bar{X}_{(i)}$ est la valeur moyenne.

2. Deuxième étape, nous avons pris la valeur absolue de chaque écart de diagraphies et nous avons normalisé en utilisant la plage. Les valeurs d'erreur de journal $N(X_i)$ sont comprises entre [0, 1] pour classer les faciès (Ameur-Zaimeche et al., 2014bc).

$$N(X_i) = \frac{ABS[D(X_i)]}{X_{i \max} - X_{i \min}} \quad (11)$$

Où $ABS[D(X_i)]$ est la déviation des données de diagraphies, $X_{i \max}$ et $X_{i \min}$ sont le maximum et le minimum de chaque log.

3. Calcul des erreurs moyennes des données de diagraphies pour chaque faciès ($Error(f)_i$)

$$Error(f)_i = meanN(X_i) \quad (12)$$

4. L'attribution de l'erreur minimale au faciès correspondant par un test logique.

4. Prédiction des Electrofaciès par les Réseaux de Neurones (MLPNN)

L'application de la technique intelligente MLPNN permis d'estimer les logs faciologiques non seulement dans les puits carottés et même de reconstituer les niveaux électrofaciès non carottés, et pour tester l'applicabilité de cette technique, il est nécessaire de faire une comparaison graphique et numérique sur les faciès réelles vs les électrofaciès.

4.1. Comparaison qualitative entre faciès carotte et électrofaciès MLPNN

La présentation graphique des résultats de la reconstitution a été faite à l'aide de la visualisation des logs des faciès de prédiction (commande discret) sur la station du logiciel PETREL (2010) par la licence du groupement Berkine (Sonatrach/Anadarko). Cela, a permet de réaliser une comparaison entre l'électrofaciès MLPNN et celui des faciès réelles observés sur les carottes des trois puits (A, B, et C). Nous avons également visualisé les électrofaciès de prédiction des puits non-carottés et les diagraphies nucléaires utilisés comme variables explicatifs dans le model MLPNN pour une bonne confirmation des résultats avec les réponses diagraphiques.

- Puits P-C

Le puits C présente un intervalle non carotté (absence de carotte) de 50 m s'étale de la profondeur (3046 m – 3110 m). Ce manque a été reconstitué par la technique MLPNN, ce dernier se présente en parfaite concordance avec les déflexions diagraphiques. Une concordance entre les limites de MLPNN et les carottes a été bien remarquée dans la partie carottée (Fig. 38).

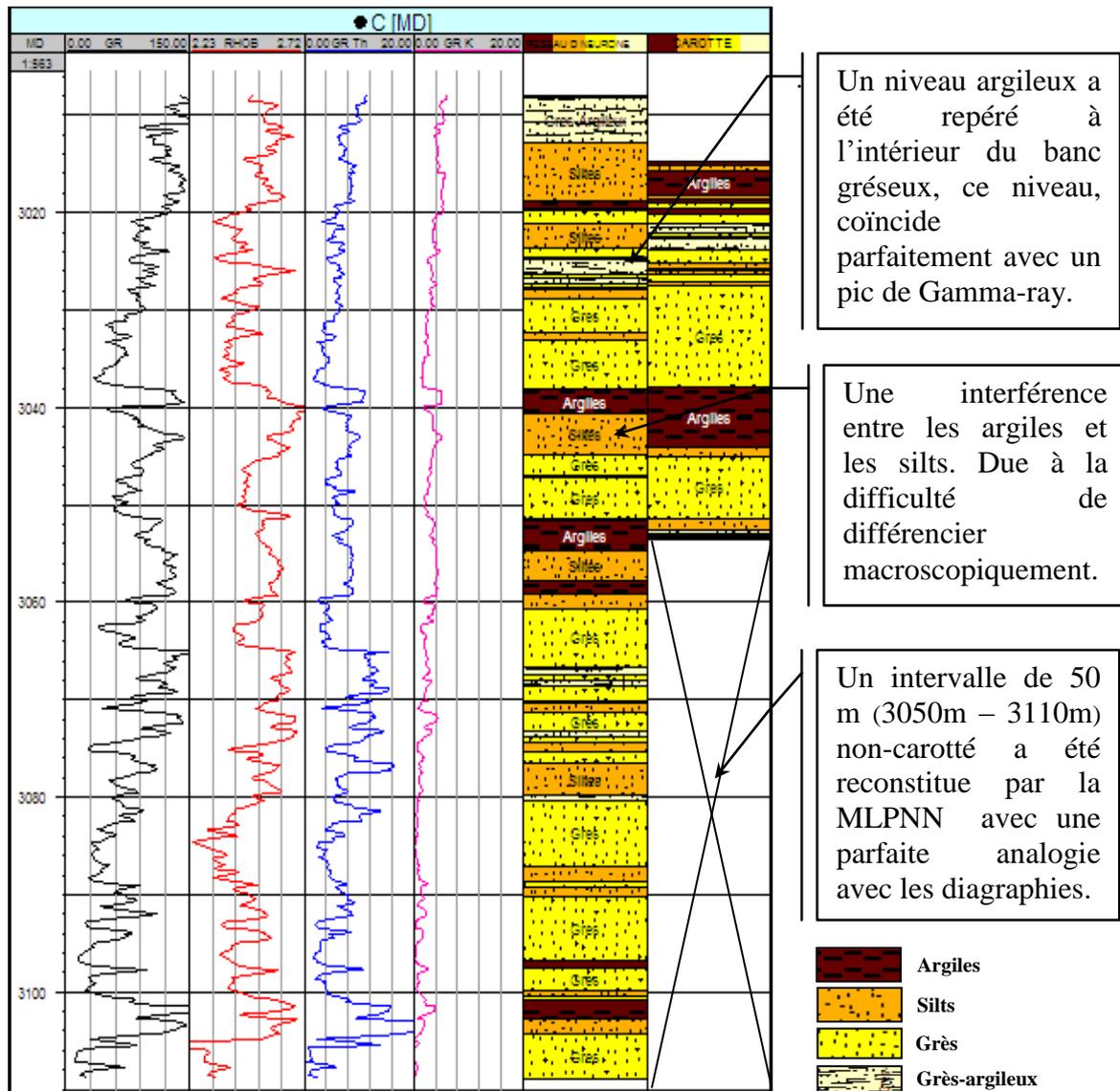


Fig. 38 - Résultat des électrofaciès MLPNN pour le puits P_C.

4.1.2. Visualisation des puits totalement non carottés

Les deux puits (P-D et P-E) sont totalement non-carottés, et ce qui se présentent un véritable test pour tester la fiabilité du modèle des MLPNN pour sourdre ce problème.

- Puits P-D

Les faciès absent ont été entièrement reconstitués et les résultats obtenus présentent que les électrofaciès estimés et les déflexions des diagraphies montrent une grande ressemblance (Fig. 39).

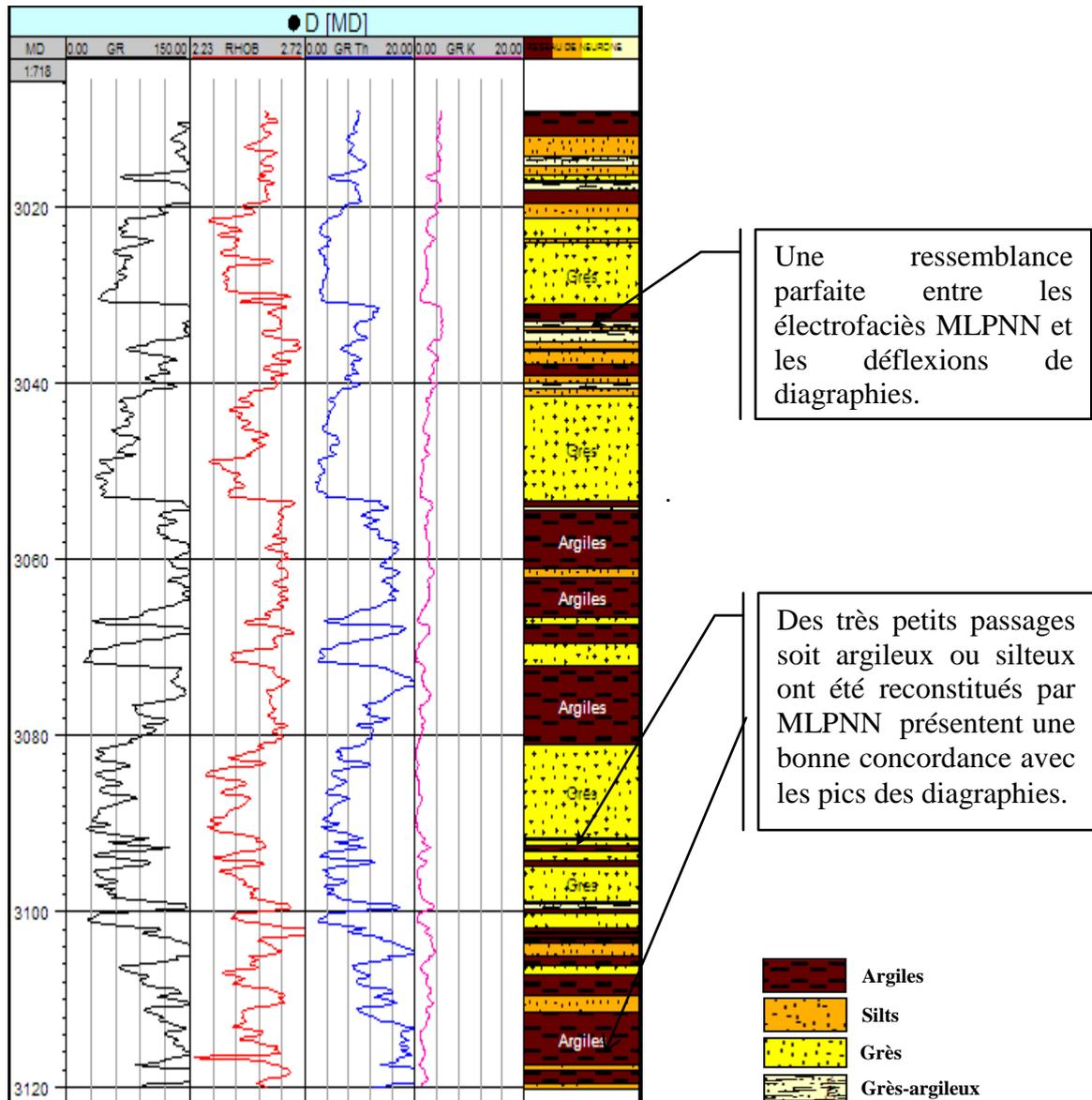


Fig. 39 - Résultat des électrofaciès MLPNN pour le puits P_D.

- Puits P-E

Ce puits n'est pas fait l'objet d'une opération de carottage, ceci il a été reconstitué intégralement par MLPNN, la visualisation graphique montre une bonne concordance entre les électrofaciès MLPNN avec les réponses des diagraphies (Fig. 40).

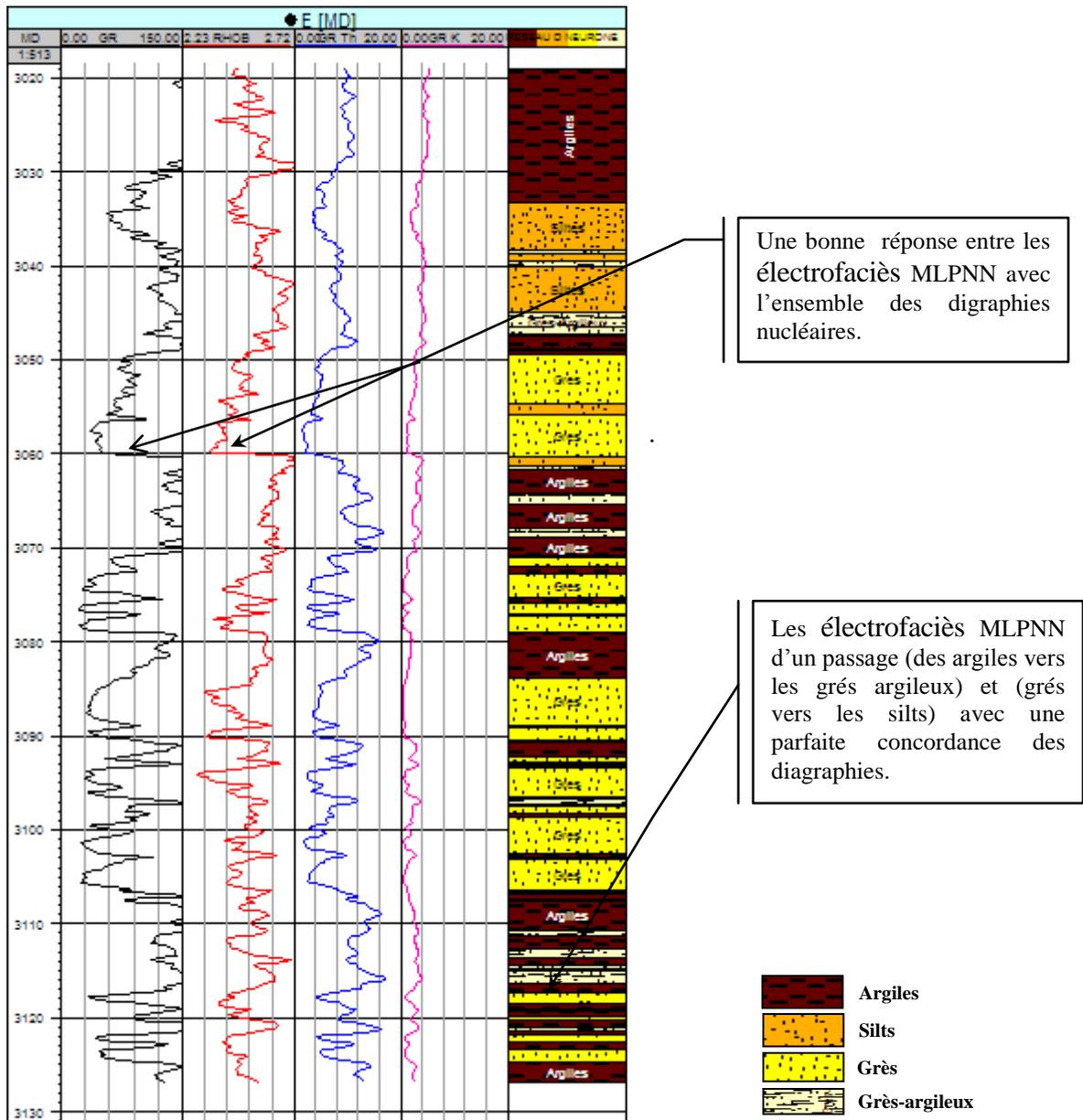


Fig. 40 - Résultat des électrofaciès MLPNN pour le puits P_E.

4.2. Comparaison quantitative entre faciès carotte et électrofaciès MLPNN

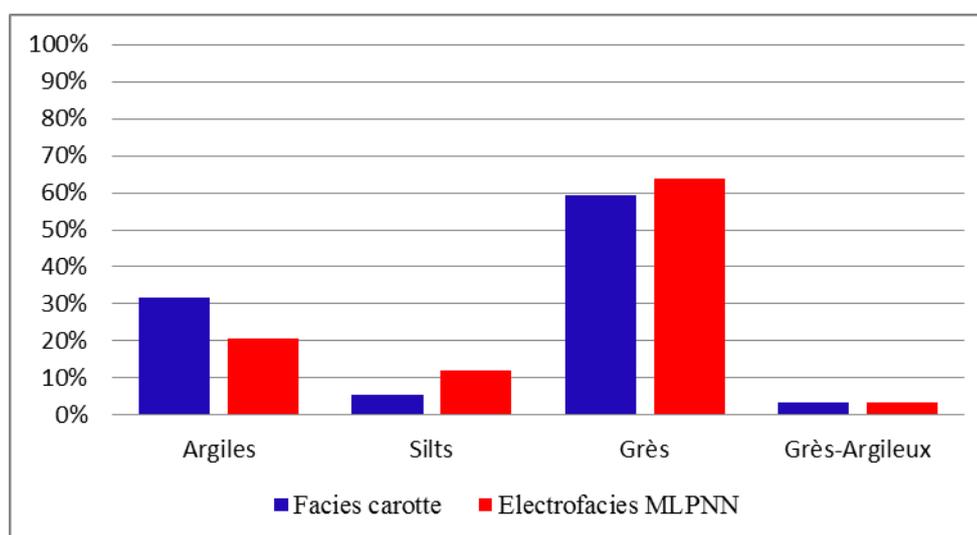
Dans cette étape on a calculé la différence pour électrofaciès (argiles, silts, grès, et grès-argileux) qui ont été déterminés par (MLPNN) et les faciès réel à partir des carottes.

4.2.1. Puits P-A

La technique MLPNN a donné des électrofaciès très proches des faciès réels des carottes. Pour les grès et les grès argileux qui présentent le faciès réservoir, il y a une différence négligeable de 5%. Pour les argiles et les silts il y a une petite interférence due à la description. (Tableau. 5).

Tableau. 5 : Comparaison des faciès carottes et électrofaciès MLPNN du puits P_A.

	Faciès	Code	Faciès Carottes	Électrofaciès MLPNN	Différence
P-A	Argiles	0	32%	21%	11%
	Silts	1	6%	12%	6%
	Grès	2	59%	64%	5%
	Grès-Argileux	3	3%	4%	1%

**Fig. 41-** Histogramme des faciès carottes et électrofaciès MLPNN puits P_A.

4.2.2. Puits P-B

On a constaté dans le puits P-B, un accroissement du pourcentage de grès-argileux et une réduction de pourcentage des argiles dans électrofaciès MLPNN, cependant pour le faciès silteux la différence est négligeable.

Tableau. 6 : Comparaison des faciès carottes et électrofaciès MLPNN du puits P_B.

	Faciès	Code	Faciès Carottes	Électrofaciès MLPNN	Différence
P-B	Argiles	0	21%	16%	6%
	Silts	1	27%	26%	1%
	Grès	2	48%	49%	2%
	Grès-Argileux	3	5%	8%	3%

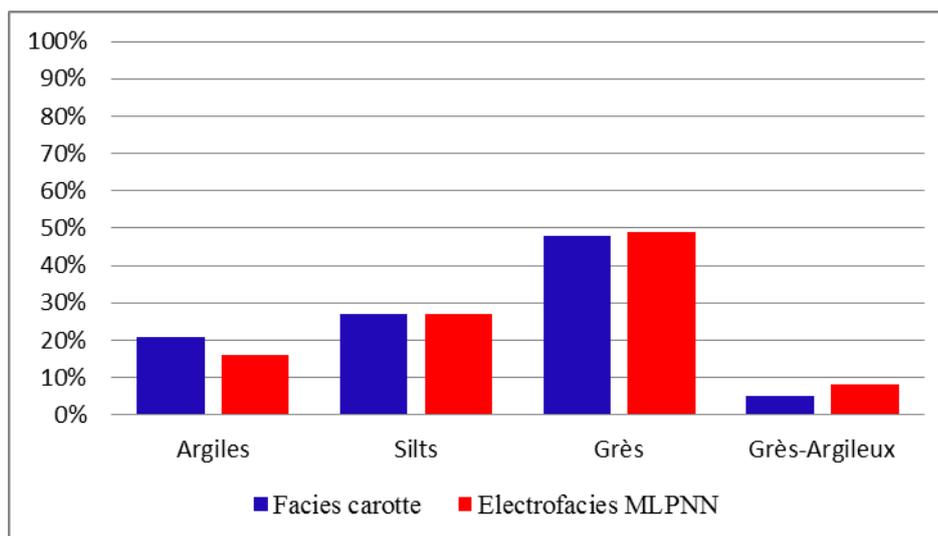


Fig. 42 - Histogramme des faciès carottes et électrofaciès MLPNN puits P_B.

4.2.3. Puits P-C

Pour le puits P-C la différence entre les faciès carottes et électrofaciès MLPNN est importante, surtout pour les argiles et les silts (10 et 11%). On a aussi une petite différence entre les grès (5%) et le grès argileux (4%).

Tableau. 7 : Comparaison des faciès carottes et électrofaciès MLPNN du puits P_C.

	Faciès	Code	Faciès Carottes	Électrofaciès MLPNN	Différence
P-C	Argiles	0	27%	17%	10%
	Silts	1	11%	22%	11%
	Grès	2	55%	50%	5%
	Grès-Argileux	3	7%	11%	4%

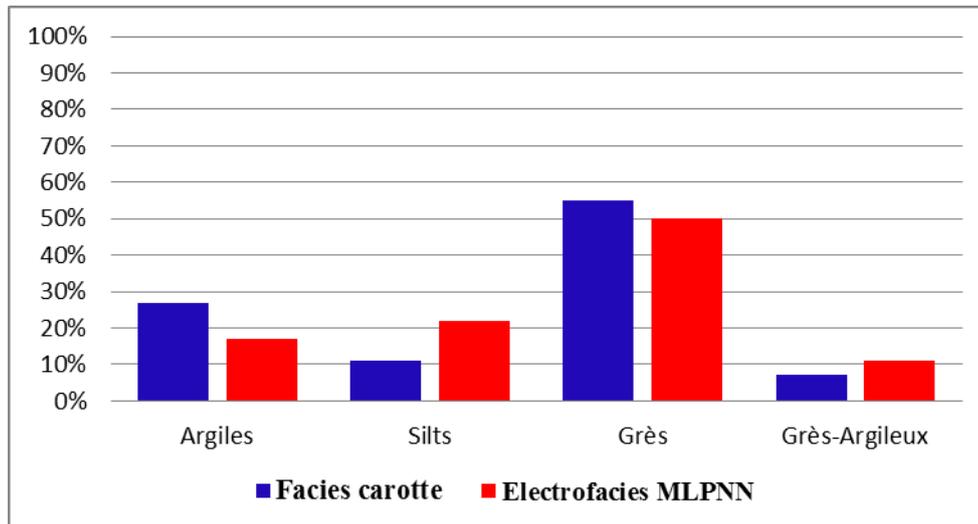


Fig. 43 - Histogramme des faciès carottes et électrofaciès MLPNN puits P_C.

4.3. Synthèse des résultats de comparaison les faciès carotte et les électrofaciès MLPNN de l'ensemble des puits A, B et C

D'après la visualisation graphique et numérique des faciès réels sur les carottes et les électrofaciès reconstitués par la technique MLPNN dans l'ensemble des puits carottés, on a constaté les observations suivantes : pour les argiles et les grès, la différence entre MLPNN et la réalité est négligeable et pour les silts et les grès-argileux la différence est de 10% et 2% respectivement voir (tableau 8 et 9).

Les petites erreurs et les interférences obtenus par MLPNN ont due au plusieurs facteurs :

- Les descriptions macroscopiques des carottes ont été faites par plusieurs géologues, et il y a une marge d'erreur pour la description visuelle sa diffère d'un géologue à l'autre.
- La description des faciès est simplifiée en quatre faciès principaux;
- La reconstitution des faciès est basée sur un traitement statistique des données sur Matlab.

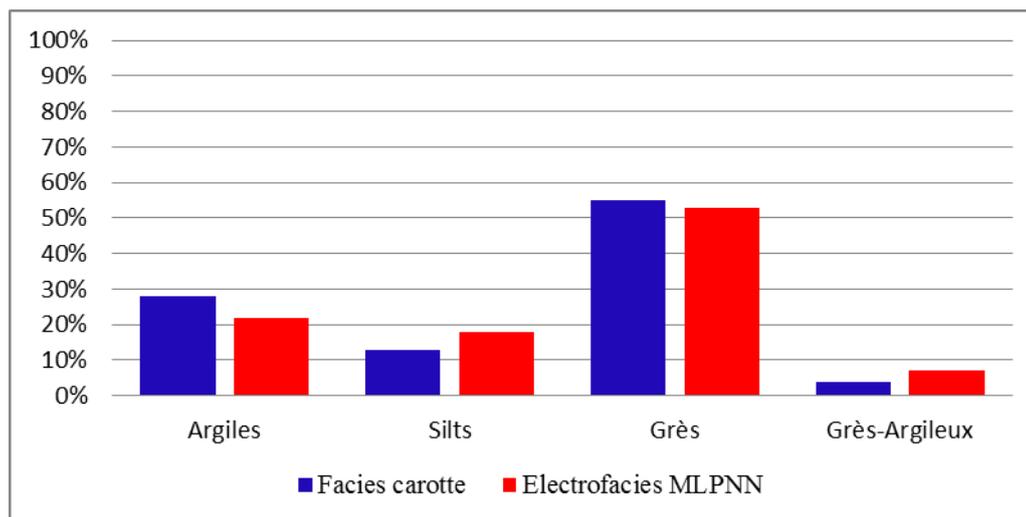
On peut discuter les petites interférences entre les argiles et les silts, qui sont liée à beaucoup de facteurs : en fonction des concentrations des éléments radioactives (U, Th, K) dans les formations (cause génétique) et le pas d'échantillonnage de l'outil de diagraphies (problème technique) et la possibilité de l'existence d'erreurs dans les descriptions (cause humain).

Tableau. 8: Statistique des faciès carottes et électrofaciès MLPNN des puits A,B,C.

Faciès	Code	Faciès carottes	Faciès MLPNN
		Nbr d'échantillons	Nbr d'échantillons
Argiles	0	344	271
Silts	1	159	245
Grès	2	678	634
Grès-Argileux	3	55	86
		1236	1236

Tableau. 9 : Comparaison des faciès carottes et électrofaciès MLPNN des puits A,B,C.

Faciès	Code	Faciès Carottes	Faciès MLPNN	Différence
Argiles	0	28%	22%	6%
Silts	1	13%	18%	5%
Grès	2	55%	53%	2%
Grès-Argileux	3	4%	7%	3%

**Fig. 44 -** Histogramme des faciès carottes et les électrofaciès MLPNN, puits A, B, C.

5. Prédiction des Electrofaciès par la Méthode de Cluster Analysis

On a pris cette méthode qui est une large application dans la reconstitution des électrofaciès, comme technique de comparaison pour tester la fiabilité avec la méthode précédente (MLPNN).

5.1. Comparaison qualitative entre faciès carotte et électrofaciès cluster analysis

5.1.1. Comparaison graphique des puits carottés

- Puits P-A

La reconstitution des électrofaciès par la technique Cluster analysis a donné pratiquement les mêmes faciès réels. Cependant pour les argiles et les silts qui il y a une interférence dans la profondeur (3000-3010 m) et deux passages (3028-3030 m) (Fig. 45).

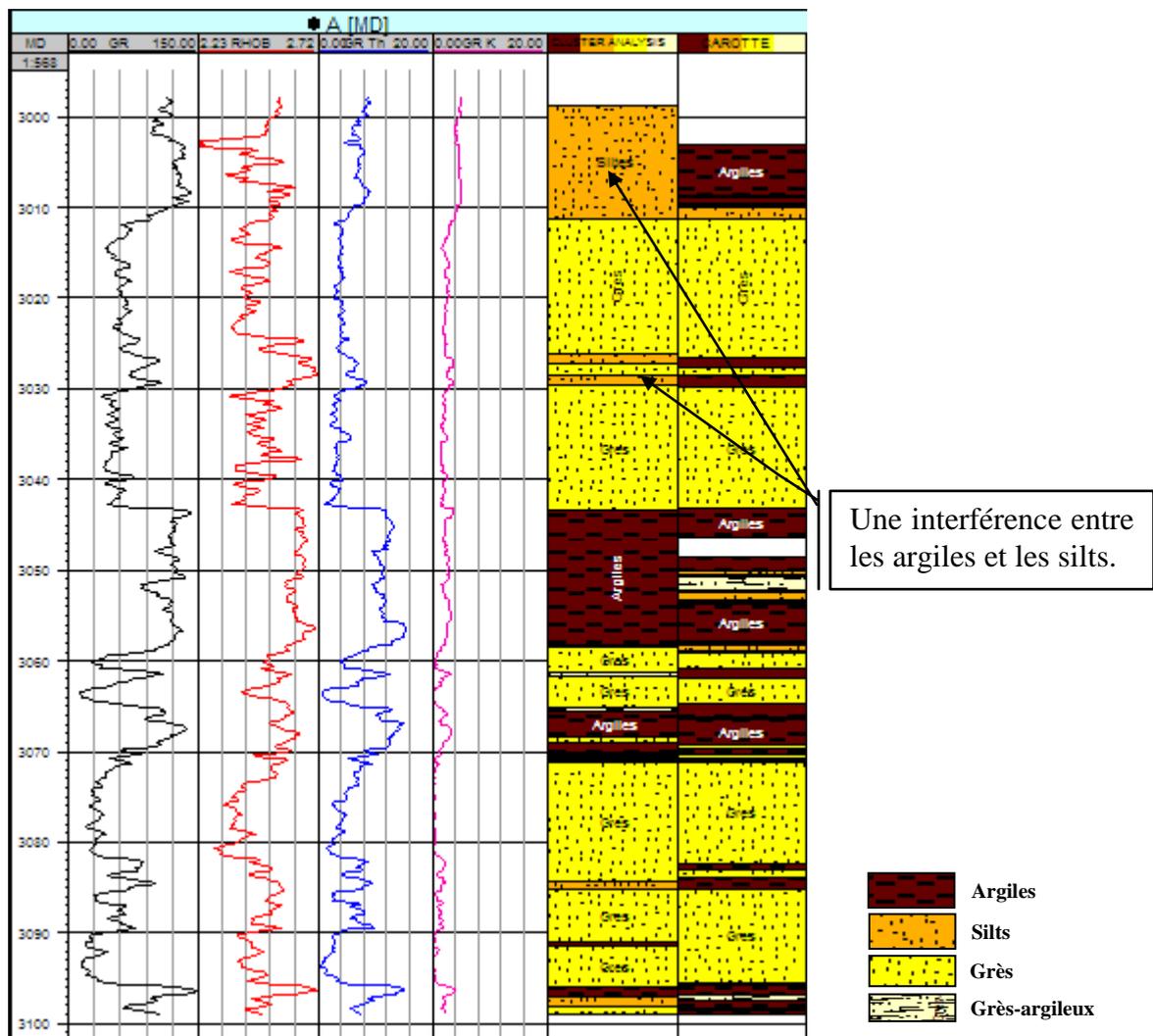


Fig. 45 - Résultat des électrofaciès cluster analysis pour le puits P_A.

- Puits P-B

Dans ce puits P-B qui est partiellement non carotté à cause d'un problème technique, on a remarqué une bonne correspondance entre les électrofaciès Cluster analysis d'une part et les faciès des carottes et les réponses diagraphiques d'autre part. On a pu de pallier de 33 m des électrofaciès dans l'intervalle non carottée (3073m et 3110 m). Il y a quelques interférences entre les argiles et les silts dans la profondeur (3020-3055 m) (Fig. 46).

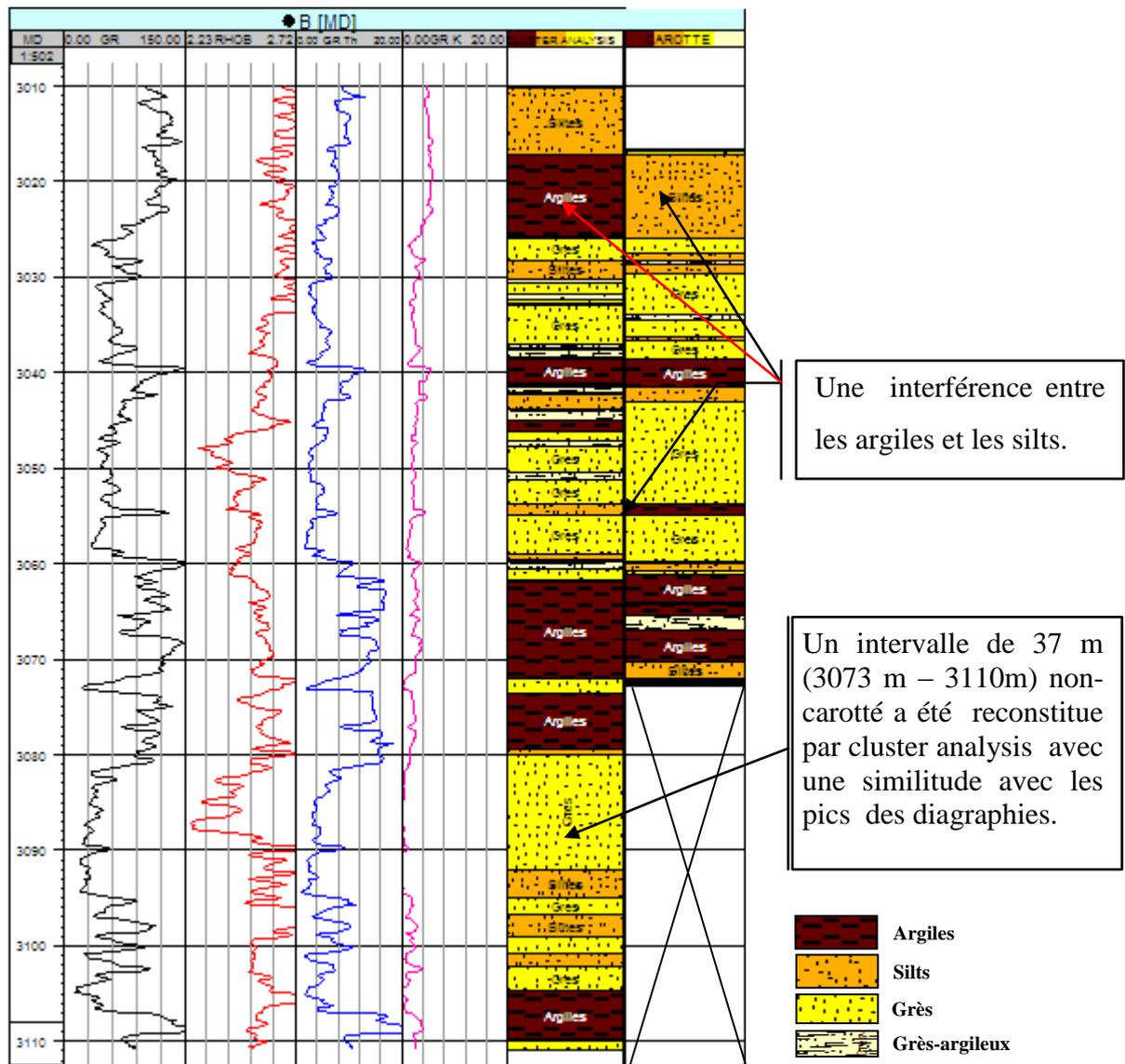


Fig. 46 - Résultat des électrofaciès cluster analysis pour le puits P_B.

- Puits P-C

Ce puits présent un intervalle non carotté (absence de carotte) de 50 m s'étale de la profondeur (3050 m – 3110 m). Le manque a été rétabli par la méthode de cluster analysis, ce dernier se présente une concordance avec les pics des diagraphiques. Une concordance entre les limites de cluster analysis et les carottes a été remarquée dans la partie carottée (Fig. 47).

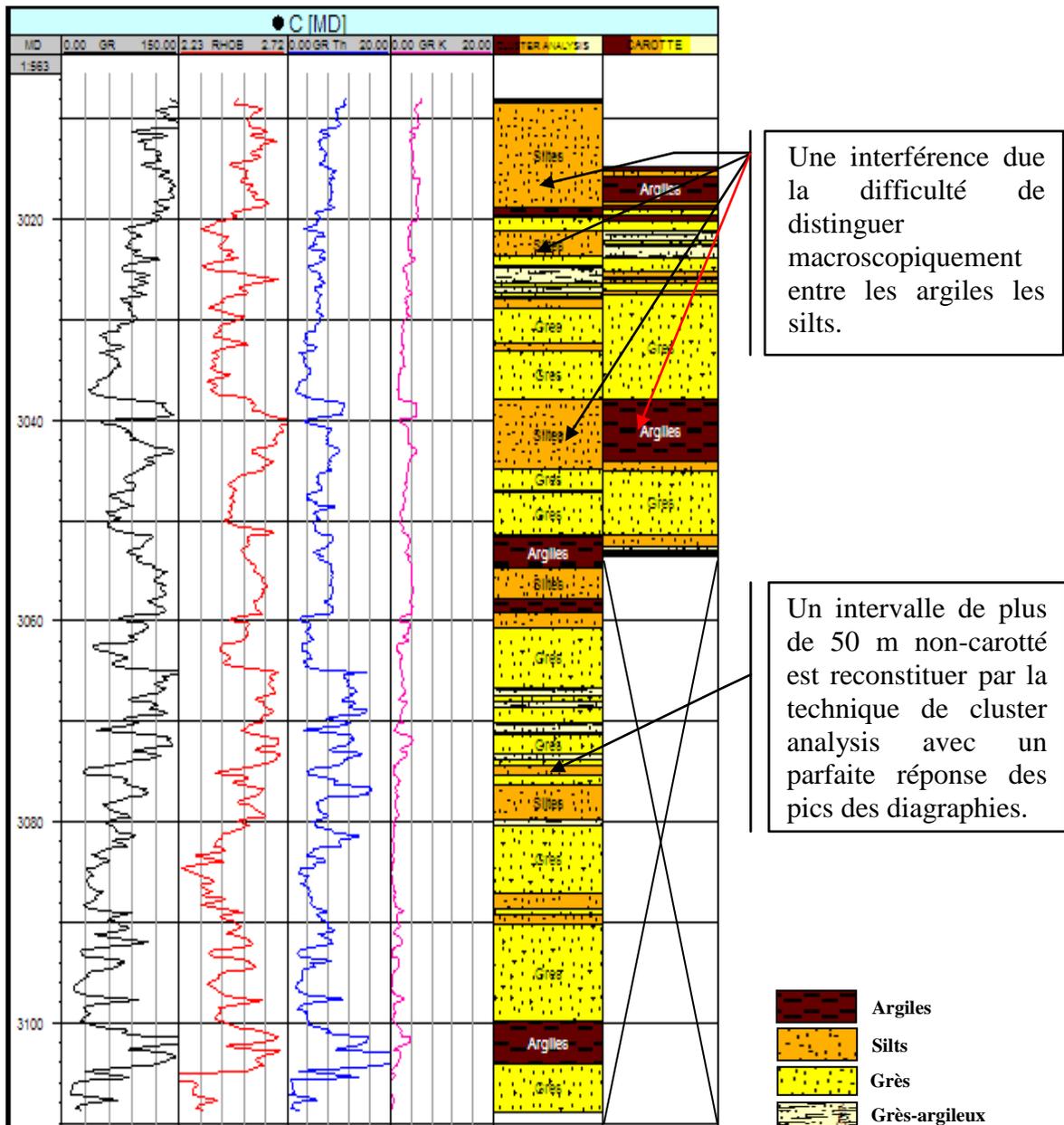


Fig. 47 - Résultat des électrofaciès cluster analysis pour le puits P_C.

5.1.2. Visualisation des puits totalement non carottés

- Puits P-D

Les lithofaciès absent ont été intégralement reconstitués et les résultats obtenus présentent que les électrofaciès estimés cluster analysis et les pics des diagraphies montrent une grande similitude surtout les logs (Gamma-ray et densité) (Fig. 48).

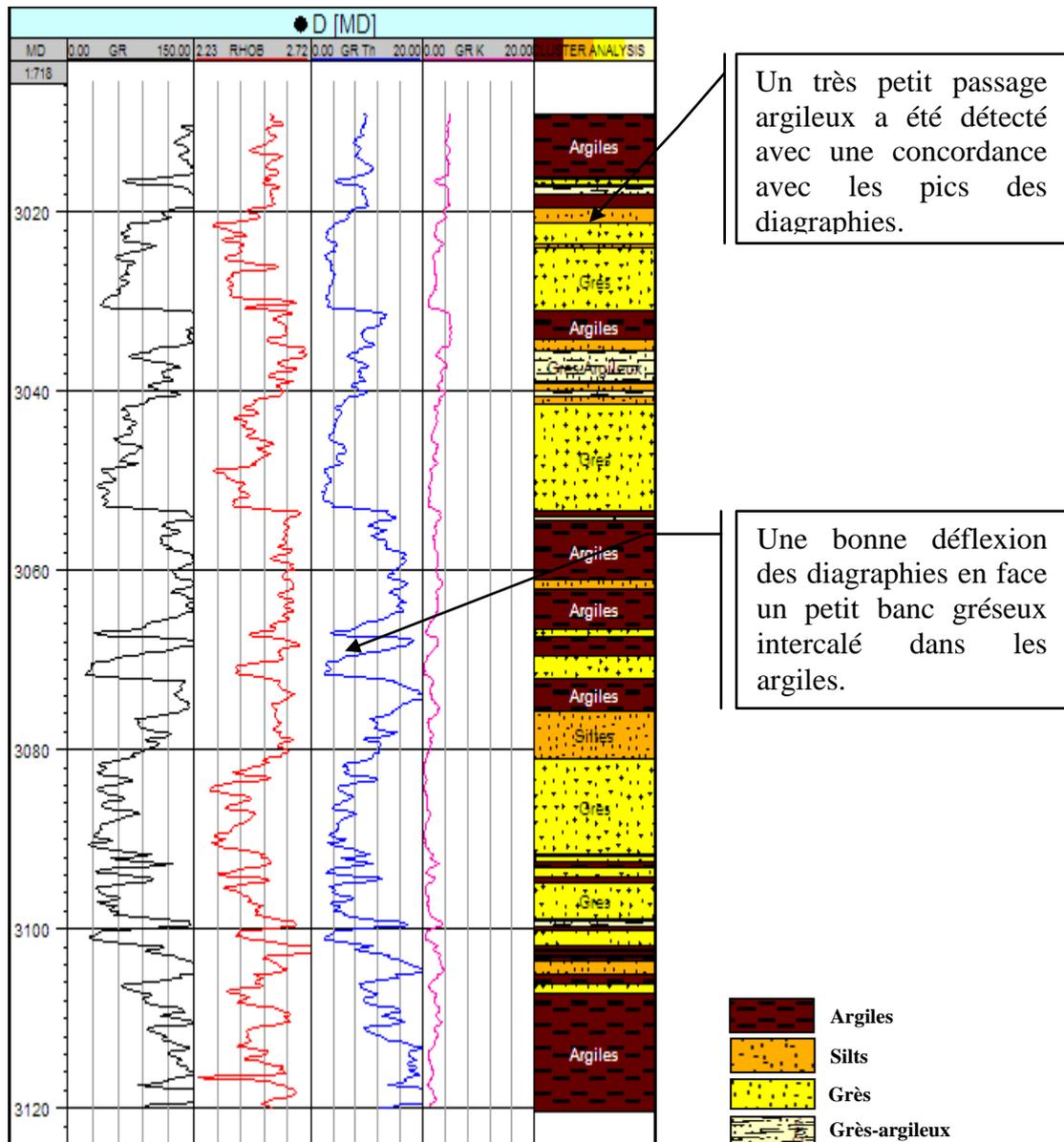


Fig. 48 - Résultat des électrofaciès cluster analysis pour le puits P_D.

- Puits P-E

Ce puits comme déjà dit précédemment, il n'a pas été l'objet de l'opération de carottage, mais il est entièrement reconstruit par la méthode de cluster analysis, et les électrofaciès cluster analysis aussi présente une bonne concordance avec les de pics des diagraphies (Fig. 49).

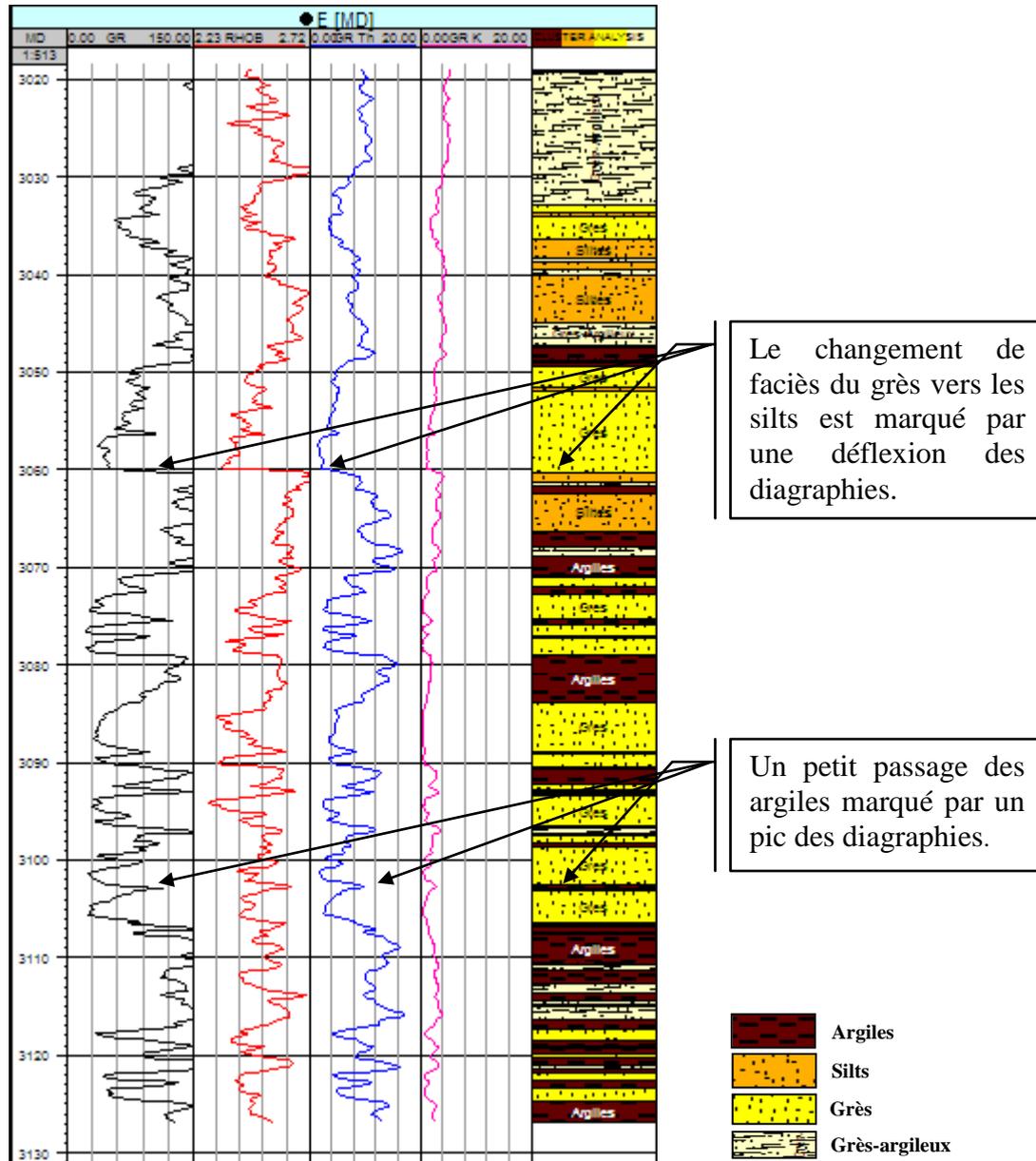


Fig. 49 - Résultat des électrofaciès cluster analysis pour le puits P_E.

5.2. Comparaison quantitative entre faciès carottes et électrofaciès cluster analysis

On va voir la différence quantitative en terme de chiffre entre le pourcentage des faciès reconstitués par la méthode de cluster analysis) pour chaque puits carotté avec les faciès réels sur carottés. Les résultats ont été représentés sous forme des tableaux et des histogrammes.

5.2.1. Puits P-A

Les résultats du cluster analysis sont relativement différentes par rapport la description réelle sur les carottes. Il y a une augmentation du pourcentage des silts et une diminution dans des argiles; pour le grès il y a une différence de 7% ; pour les grès argileux la différence est négligeable (Fig. 50).

Tableau. 10 : Comparaison des faciès carottes et électrofaciès cluster analysis, puits P_A.

	Faciès	Code	Faciès Carottes	Electrofaciès cluster Analysis	Différence
P-A	Argiles	0	32%	23%	8%
	Silts	1	6%	22%	16%
	Grès	2	59%	52%	7%
	Grès-Argileux	3	3%	3%	0%

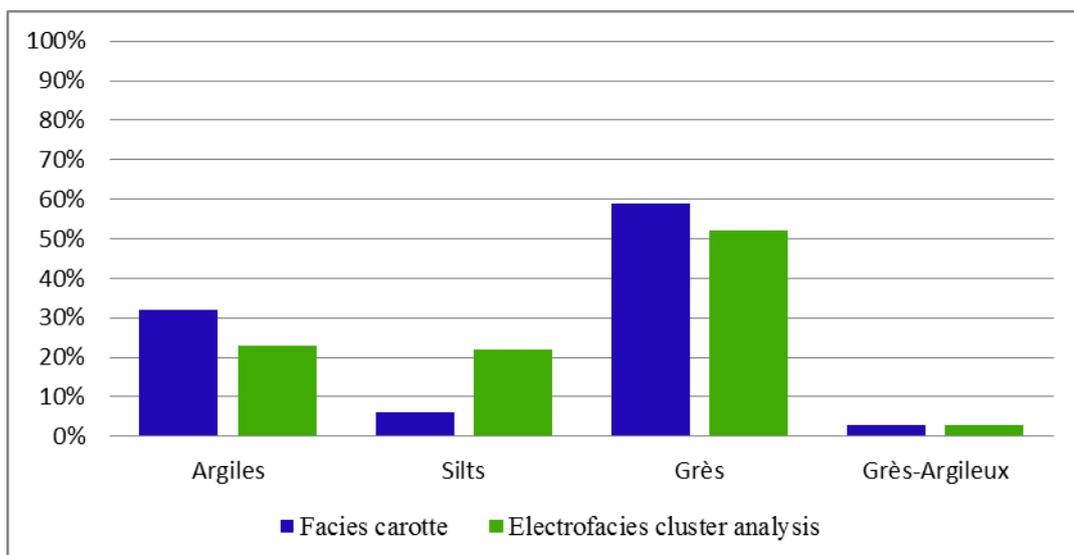


Fig. 50 - Histogramme des faciès carottes et électrofaciès cluster analysis puits P_A.

5.2.2. Puits P-B

Pour ce puits on a constaté, que le pourcentage des grès a été diminué par la technique de cluster analysis comparant avec le faciès réels par une différence de 7%, parallèlement, il y a une augmentation du pourcentage de grès-argileux et les argiles (Fig. 51) (Tableau. 11) :

Tableau. 11 : Comparaison des faciès carottes et électrofaciès cluster analysis, puits P_B

Faciès	Code	Faciès Carottes	Electrofaciès cluster analysis	Différence
Argiles	0	21%	30%	9%
P-B Silts	1	27%	20%	7%
Grès	2	47%	40%	7%
Grès-Argileux	3	5%	10%	5%

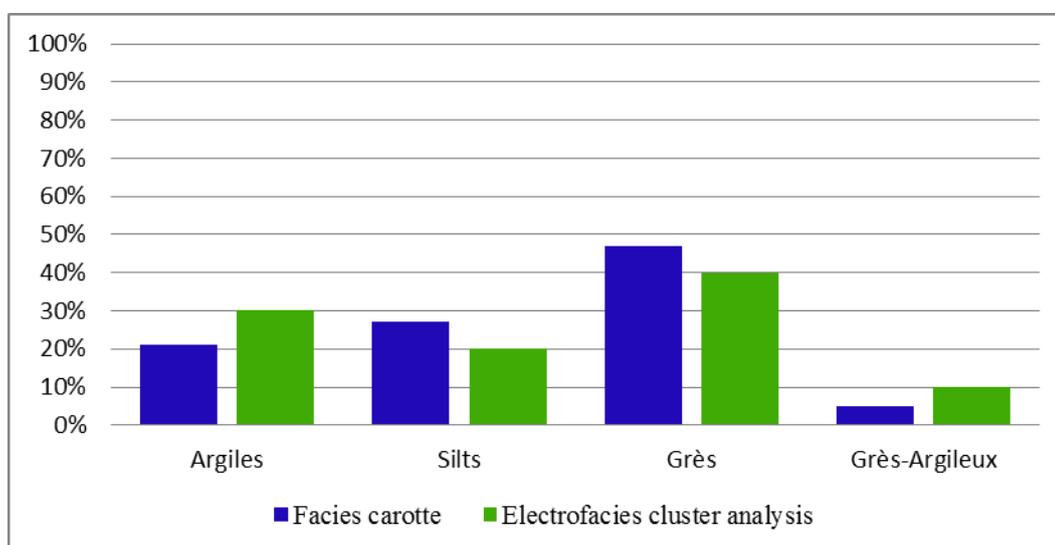


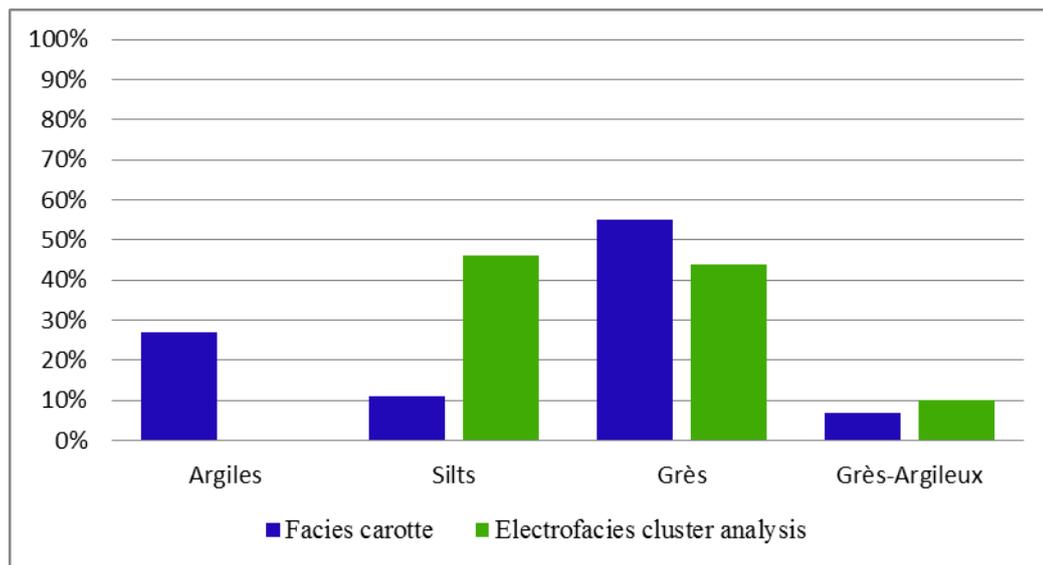
Fig. 51 - Histogramme des faciès carottes et électrofaciès cluster analysis puits P_B.

5.2.3. Puits P-C

La comparaison entre les électrofaciès et les faciès réels dans ce puits est très remarquable par rapport les deux autres puits (A et B), avec une différence (27%) pour les argiles et (35%) pour les silts 35%, et une différence entre les grès (11%) et le grès argileux (3%), (Fig. 52), (Tableau. 12).

Tableau. 12 : Comparaison des faciès carottes et électrofaciès cluster analysis, puits P_C.

Faciès	Code	Faciès Carottes	Electrofaciès cluster analysis	Différence	
P-C	Argiles	0	27%	0%	27%
	Silts	1	11%	46%	35%
	Grès	2	55%	44%	-11%
	Grès-Argileux	3	07%	10%	3%

**Fig. 52 -** Histogramme des faciès carottes et électrofaciès cluster analysis puits P_C.

5.3. Synthèse des résultats de comparaison faciès carotte et électrofaciès cluster analysis de l'ensemble des puits A, B et C

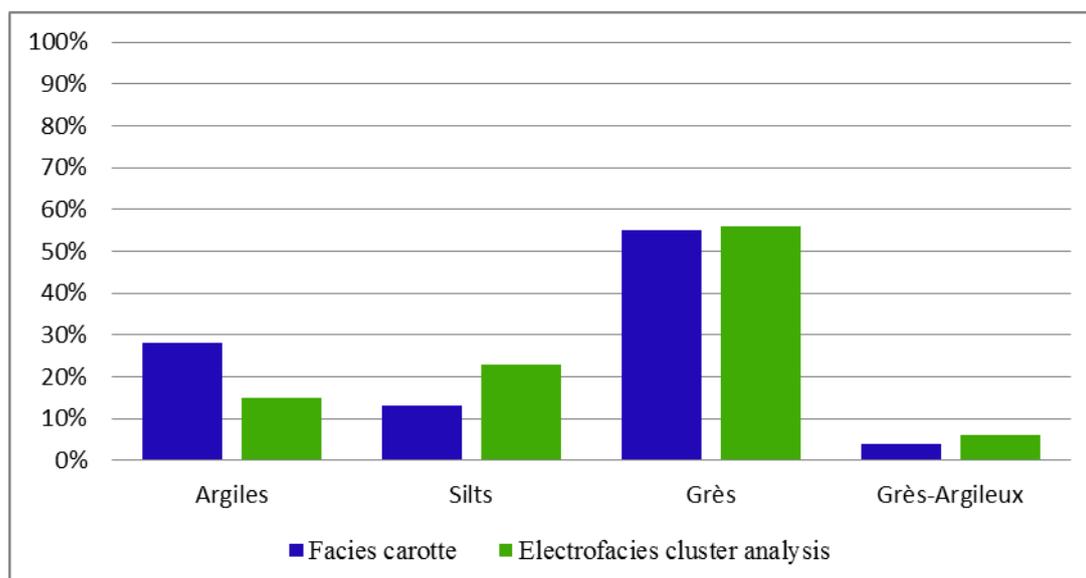
D'après la visualisation graphique et numérique des électrofaciès reconstitués par le technique cluster analysis avec les faciès réels sur les carottes dans l'ensemble des puits carottés, on a remarqué une légère différence pour le grès et les gréseux argileux qui présentent le faciès réservoir avec une différence de (1%), et pour les argiles et les silts il y a une différence remarquable de (10% et 13%) respectivement voir (tableau 13 et 14).

Tableau. 13 : Statistique des faciès carottes et électrofaciès cluster analysis

Faciès	Code	Faciès carottes	cluster analysis
		Nbr d'échantillons	Nbr d'échantillons
Argiles	0	344	221
Silts	1	159	298
Grès	2	678	606
Grès-Argileux	3	55	111
		1236	1236

Tableau. 14 : Comparaison des faciès carottes et électrofaciès cluster analysis dans les puits (A,B,C).

Faciès	Code	Faciès Carottes	Electrofaciès custer analysis	Différence
Argiles	0	28%	15%	13%
Silts	1	13%	23%	10%
Grès	2	55%	56%	1%
Grès-Argileux	3	4%	6%	2%

**Fig. 53 -** Histogramme des faciès carotte et les électrofaciès cluster analysis dans les puits A,B,C

6. Comparaison Entre les Résultats des Electrofaciès MLPNN et de Cluster Analysis

Cette étape est très importante pour tester la fiabilité des deux méthodes appliqués dans notre étude pour résoudre le problème du manque des faciès dans les puits partiellement ou totalement non carottés d'où la nécessité de pallier avec précision les lacunes en vue une bonne caractérisation du réservoir et le bon choix de la méthode qui donne le moindre des erreurs, et pour cela une visualisation graphique et le traitement numériques des résultats des électrofaciès de (MLPNN et cluster analysis) a été faite sur l'ensemble des puits carottés (P-A, P-B et P-C).

Tableau. 15: Comparaison des faciès carottes et MLPNN ; cluster analysis) électrofaciès P-A.

	Faciès	Code	Faciès Carottes	Électrofaciès MLPNN	Électrofaciès cluster analysis
P-A	Argiles	0	32%	21%	23%
	Silts	1	06%	12%	22%
	Grès	2	59%	64%	52%
	Grès-Argileux	3	03%	04%	03%

Tableau. 16: Comparaison des faciès carottes et (MLPNN ; cluster analysis) électrofaciès P-B.

	Faciès	Code	Faciès Carottes	Électrofaciès MLPNN	Électrofaciès cluster analysis
P-B	Argiles	0	21%	16%	30%
	Silts	1	27%	27%	20%
	Grès	2	48%	49%	40%
	Grès-Argileux	3	05%	08%	10%

Tableau. 17: Comparaison des faciès carottes et (MLPNN ; cluster analysis) électrofaciès P-C.

	Faciès	Code	Faciès Carottes	Électrofaciès MLPNN	Électrofaciès cluster analysis
P-C	Argiles	0	27%	20%	00%
	Silts	1	11%	26%	46%
	Grès	2	55%	42%	44%
	Grès-Argileux	3	07%	12%	10%

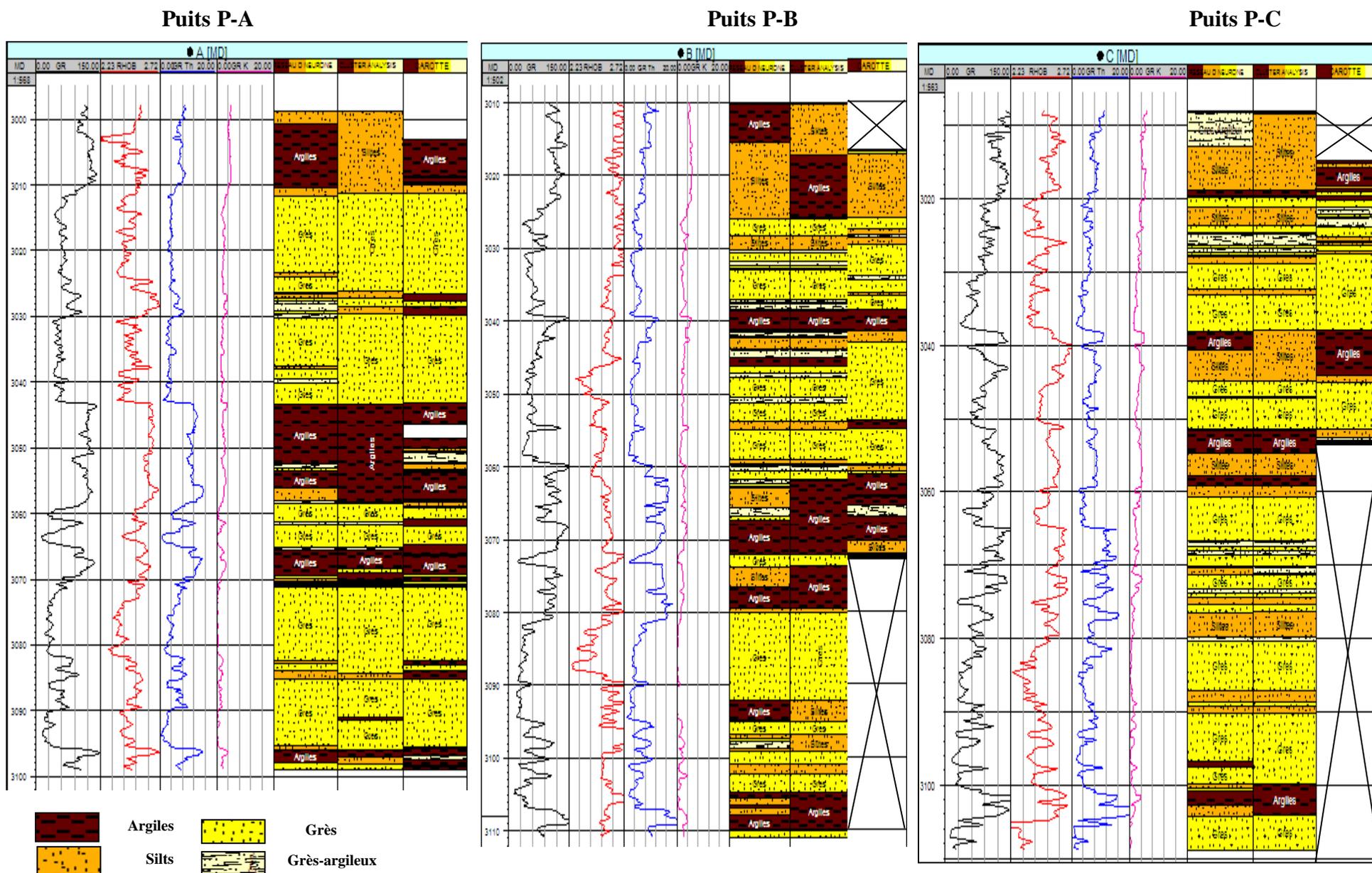


Fig. 54– Comparaison des lithofaciés de carotte avec les résultats (MLPNN et cluster analysis) électrofaciés des puits (A, B, C)

En basant sur les tableaux (15, 16, 17) de comparaison et la figure 54 qui englobent les faciès réels des trois puits (A, B, C) avec les électrofaciès prédites par les deux méthodes, il apparaît clairement que la technique MLPNN a donné des meilleurs résultats de reconstitution des faciès non carottés face à la deuxième méthode cluster analysis.

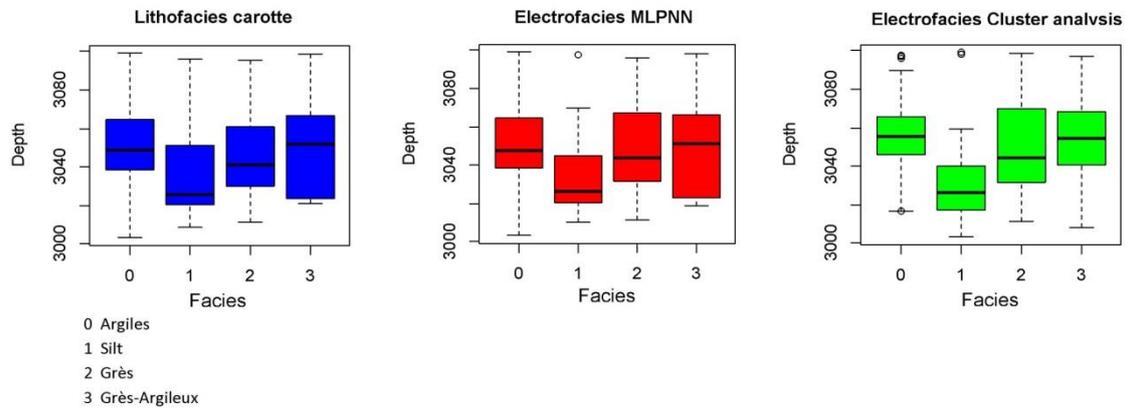


Fig. 55– Boite à moustache des lithofaciès de carotte avec les résultats (MLPNN et Cluster analysis) électrofaciès des puits carottés (A, B, C)

Les boîtes à moustaches ont été établies pour une meilleure comparaison des deux techniques de prédiction avec les carottes réel. La précision de l'approche MLPNN a été démontrée en outre, à partir des boîtes à moustaches montrant une correspondance complète entre les lithofaciès de carotte observés et les électrofaciès MLPNN, par rapport les boîtes à moustaches de la méthode de cluster analysis (Fig.55).

Conclusion

Dans le présent chapitre on a investigué l'application du MLPNN et la méthode de cluster analysis pour prédire les lithofaciès non carottés dans le réservoir hétérogène (TAGI), du gisement de Sif Fatima. L'application de ces techniques aux données de diagraphie nucléaires qui sont combinées avec les lithofaciès observés dans des puits carottés. Selon les résultats obtenus, nous pouvons conclure ce qui suit:

1. Les modèles MLPNN et cluster analysis doivent être définis sur des données de diagraphies nucléaires en entrée (Input): gamma ray, densité, thorium et potassium qui représentent avec précision un proxy des lithofaciès tandis que d'autres (par exemple, résistivité et sonique) peuvent inclure des interférences en raison de leur sensibilité au fluide interstitiel et la fracturation de la roche.

2. Le modèle de MLPNN avec quatre entrées, quatre neurones dans la couche cachée et les lithofaciès en sortie, est la meilleure architecture pour prédire les lithofaciès non carottés. Cette architecture (4-4-4) est réalisée avec l'algorithme d'apprentissage de rétropropagation basé sur Levenberg-Marquardt et la fonction de transfert sigmoïde où le modèle MLPNN atteint les meilleures performances lors de la formation, de la validation et du test (corrélation R^2 : de 0,78 à 0,92; RMSE: de 0,39 à 0,55; MAE: de 0,23 à 0,30).
3. La comparaison entre le MLPNN et les modèles de cluster analysis a prouvé que le MLPNN est le meilleur pour prédire les électrofaciès. La comparaison entre les noyaux réels et les électrofaciès simulés indique que le MLPNN récupère bien le faciès de grès et de siltstone tandis que de petites interférences dans l'argile, le grès-argilo-argileux peuvent se produire. Au contraire, les interférences dans cluster analysis sont plus prononcées avec des performances plus faibles ($R^2= 0,68$, RMSE = 1,04 et MAE = 0,54).
4. L'application de ces méthodes a permis de reconstruire 333 m de faciès non carottés dans le champ pétrolifère de Sif Fatima mettant en évidence une approche intelligente, rapide et totalement économique pour les études souterraines. Ils aident les compagnies pétrolières à réduire les coûts des opérations de carottage.

Chapitre IV

Prédiction de la perméabilité et la porosité

Chapitre IV. Prédiction de la Perméabilité et la Porosité

Introduction

La caractérisation de transport et de stockage des hydrocarbures, y compris la porosité et la perméabilité sont cruciales dans le plan d'évaluation et de développement des réservoirs pétroliers. Une estimation précise de ses paramètres, en particulier la perméabilité, est essentielle pour l'amélioration de la récupération de pétrole / gaz, la sélection du régime de production, l'optimisation du placement des puits, ainsi la porosité pour l'estimation des réserves en place, alors une sous ou sur estimation de ce dernier cause des grandes problèmes dans la rentabilité du gisement.

Le meilleur moyen d'avoir une bonne caractérisation de ses propriétés est la mesure direct sur carottes au laboratoire, cependant les carottes n'ont pas toujours disponibles à cause des problèmes techniques et géologiques rencontrés durant l'opération du carottage (Ameur-Zaimeche et al., 2020), Bien que le processus de carottage soit très coûteux et demande beaucoup de temps, il peut influencer le résultat en n'ayant pas assez de carottes à évaluer (Newman & Martinn, 1977; Byrnes,1994). Cela peut entraîner une incertitude très élevée accompagnée des données géologiques extrapolées ou estimées. Alors il faut chercher un modèle efficace, rapide et compétent dans l'estimation de la porosité et la perméabilité.

Diverses formules ont été suggérées dans la littérature pour la prédiction de perméabilité et de porosité, cependant, la plupart de ces corrélations ne sont appropriées que pour l'estimation de la perméabilité/porosité grès, sables non consolidés et milieux poreux relativement homogènes (Newman & Martin, 1977 ; Thomas & Pugh, 1987). C'est difficile de développer une corrélation étendue de porosité / perméabilité pour les réservoirs très hétérogènes, y compris les roches fluviatile type argileux gréseux à porosité triple ou double. La plupart des formules proposées incluent des données statiques sur les propriétés physiques des roches, tandis que la perméabilité est une variable dynamique et son estimation d'une plus grande confiance nécessite un historique de production instantané (Dullien, 1992). Cependant, avoir des données dynamiques peut nécessiter plus de temps et également des coûts plus élevés (Byrnes, 1994).

Traditionnellement, les relations empiriques et des techniques de régression non paramétrique ont été utilisées à pour résoudre ses problèmes (Wendt et al., 1985; Amaefule et al., 1993; Xue et al., 1996; Mohaghegh et al., 1997; Datta-Gupta et al., 1999).

Au cours des deux dernières décennies, de nombreuses tentatives et méthodes ont été proposées pour appliquer des algorithmes d'intelligence artificielle pour générer des mesures de porosité et de perméabilité à partir de l'enregistrement de diagraphies (Jamialahmadi et Javadpour, 2000; Lim, 2005; Mohaghegh, 2005; Wood, 2019).

Dans le présent chapitre on va appliquer et comparer les performances de trois algorithmes de l'intelligence artificielle qui sont (Support Vector machine, Random Forest, et les réseaux de neurones artificiels) sur les données de diagraphies pour prédire la porosité et la perméabilité dans le réservoir du Trias argileux Gréseux Inferieur du champ de Sif Fatima bassin de Berkine.

1 Méthodologie

1. 1. Analyses et préparation de la base de données

Un total de 332 de mesures réels de porosité et de perméabilité (POR CORE, PERM CORE) qui ont été dérivées des trois puits carottés (A, B, et C) de réservoir TAGI de la zone de Sif Fatima, avec leurs enregistrements diagraphiques : Profondeur (DEPTH), Neutron (TNPH), le Temps de Transit (DT), la Densité Apparente (RHOB) la Résistivité (AHO10, AHO90) le Gamma Ray classique (GR), Potassium (POTA), Thorium (THO), le Caliper (HCAL) ; ainsi la porosité neutron (PHIE), et la saturation en eau (S_w) comme paramètres interprétés à partir des données de diagraphies. Cette étape consiste à rassembler, combiner, structurer, organiser et voir l'homogénéité, la similarité des données afin de pouvoir les analyser dans le cadre de programmes d'informatique.

1.1.1. Statistique mono-varié

Le traitement statistique a été effectué sur l'ensemble la porosité et la perméabilité mesurés au laboratoire et tous les enregistrements diagraphies des puits carottés, les résultats obtenus du traitement statistique sont présentés dans le tableau.

Tableau. 18 : Statistiques de porosité, perméabilité et des données de diagraphies

Parametres	Min	Max	Médiane	Moyenne	Variance	Ecart-type	CV
PERM CORE	0.009	272.591	8.132	33.007	2792.796	52.847	1.601
POR CORE	2.170	19.554	13.317	12.785	12.842	3.584	0.280
DEPTH	3110.	3195	3137	3143.966	488.450	22.101	0.007
TNPH	0.097	0.260	0.162	0.163	0.001	0.027	0.165
DT	63.381	88.643	72.968	72.729	12.618	3.552	0.049
RHO	2.329	2.815	2.466	2.474	0.006	0.078	0.031
AHO10	0.415	13.988	2.542	2.935	3.069	1.752	0.597
AHO90	0.365	9.950	2.056	2.459	2.777	1.666	0.678
GRCC	5.470	119.139	51.793	53.278	441.745	21.018	0.394
POTA	0.132	5.578	2.182	2.251	1.142	1.069	0.475
THOR	0.506	16.639	4.323	4.859	5.000	2.236	0.460
HCAL	5.907	6.202	6.041	6.035	0.003	0.054	0.009
PHIE	0.000	21.500	12.5	11.8	0.3	5.2	0.439
S_w	0.151	1.000	0.640	0.650	0.067	0.259	0.399

D'après les résultats mon-variés sur l'ensemble des data, la perméabilité avec un étendu varie de Min=0.009 md Max=272.591 md, et une variance de 2792.796. Pour la porosité min=2.170 %, max= 19.554 % avec une variance de 12.842.

1.1.2. Statistiques bi-variés (matrice de corrélation)

Le calcul de la matrice de corrélation a été faite sur l'ensemble des données de diagraphies avec de porosité et de perméabilité des puits disponibles et La matrice de corrélation a montré que la porosité (POR CORE) est corrélée positivement avec PHIE (CC=0,66), TNPH= (CC=0.27), DT (CC=0.22) et négativement avec S_w (CC= -0,46), RHOB (CC=-0,66), GR (CC=-0.20), THO (CC=-0.46), HCAL (CC=-0.30).

La perméabilité (PERM CORE) a des très faibles coefficients de corrélations avec l'ensembles des autres paramètres.

NB/ La signification des coefficients de corrélation (en rouge) ont été confirmés par la formule suivante:

$$R = 2/\sqrt{(n-1)} \quad (13)$$

Tableau. 19: Matrice de corrélation (porosité et la perméabilité carotte) et les enregistrements diagraphiques.

	PHIE	S _w	TNPH	DT	RHOB	AHO10	AHO90	GR	POTA	THOR	HCAL	PERM CORE	PORE CORE	
PHIE	1	-0.56	0.19	0.27	-0.97	-0.17	-0.10	-0.31	-0.12	-0.53	-0.23	0.17	0.66	
S _w		1.00	-0.31	0.43	0.56	-0.40	-0.58	-0.18	-0.36	0.25	0.19	0.13	-0.46	
TNPH			1.00	0.75	-0.23	0.00	0.14	0.54	0.59	0.24	-0.12	0.03	0.27	
DT				1.00	-0.30	-0.05	0.15	0.60	0.67	0.19	0.05	-0.15	0.22	
RHOB					1.00	0.14	0.07	0.28	0.09	0.52	0.26	-0.15	-0.66	
AHO10						1.00	0.83	0.15	0.12	0.13	-0.09	-0.02	0.01	
AHO90							1.00	0.27	0.27	0.13	-0.08	-0.13	0.02	
GR								1.00	0.93	0.69	0.06	-0.31	-0.20	
POTA									1.00	0.37	-0.06	-0.26	-0.02	
THOR										1.00	0.19	-0.25	-0.46	
HCAL											1.00	-0.02	-0.30	
PERM CORE												1.00	0.35	
PORE CORE														1

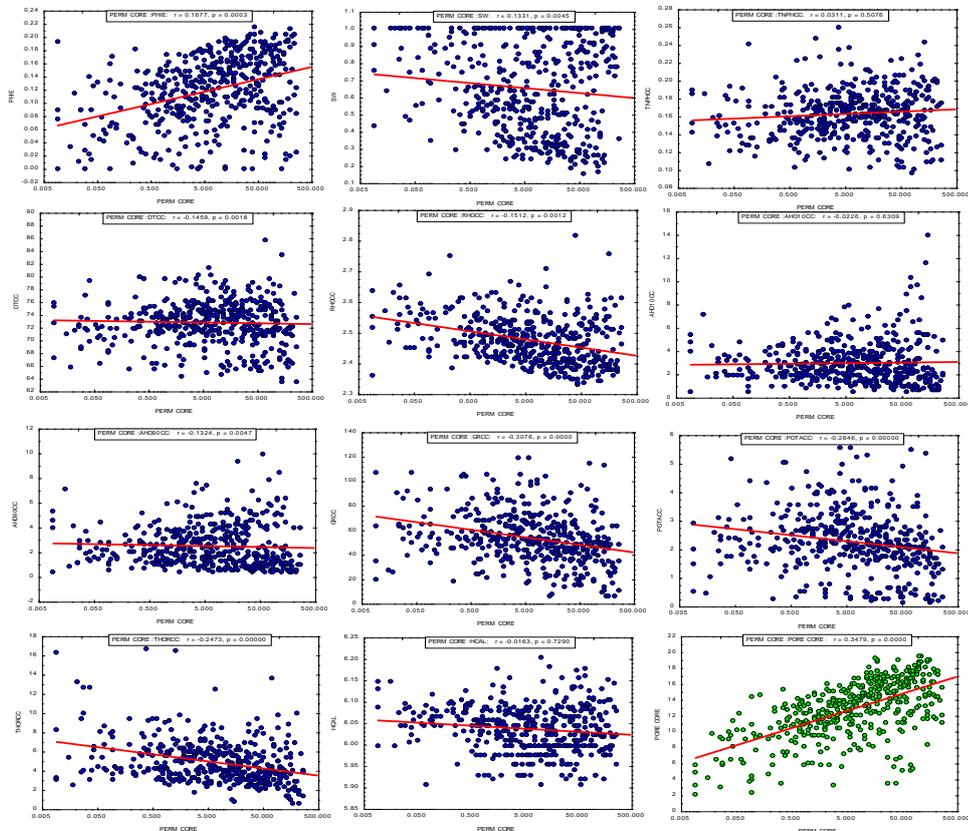


Fig. 56 - Cross-plots de la perméabilité avec les différents paramètres.

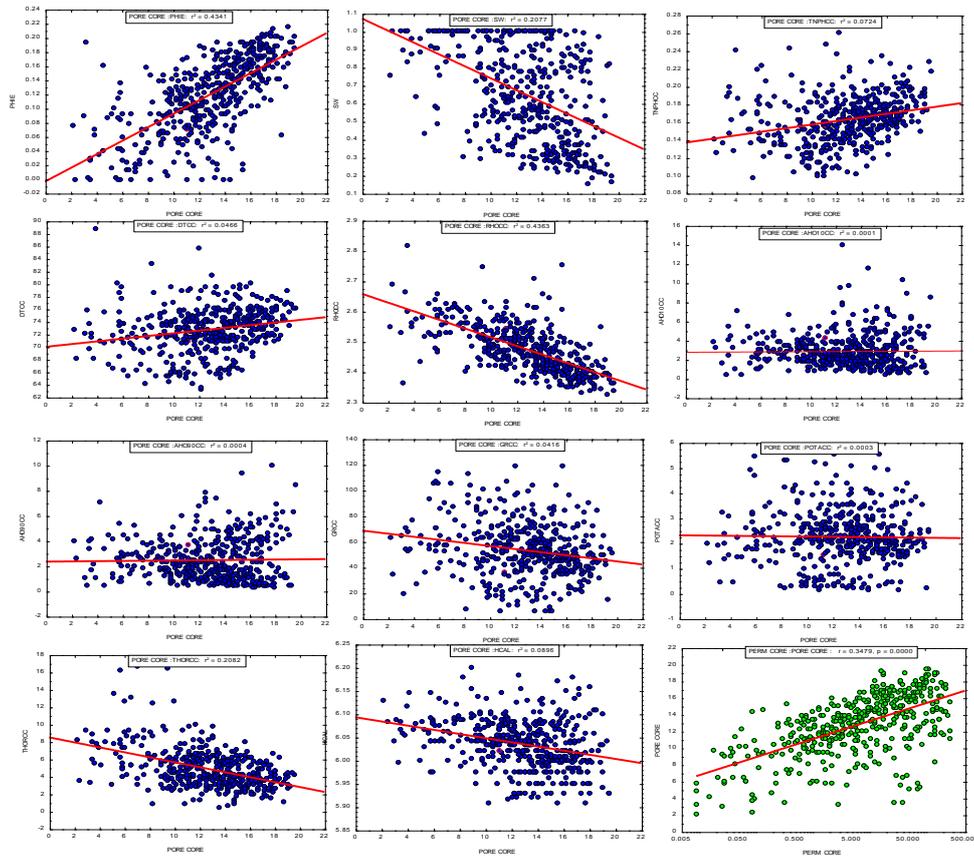


Fig. 57 - Cross-plots de la porosité avec les différents paramètres.

Pour une meilleure visualisation graphique de la porosité et la perméabilité mesurée avec les différents paramètres les cross-plots (Figure. 56.) ont été établies, pour la porosité il y a une certaine corrélation avec les paramètres surtout (PHIE, NPHIE, et ROHB) mais pour la perméabilité il y a une absence totale de corrélation significatif (Fig. 57), et même entre la porosité carotte et perméabilité carotte il y a une faible corrélation (plot coloré en vert).

1.1.3. Statistiques multi-variés Analyse en Composantes Principales (ACP)

Elle a été introduite par (Pearson, 1901), et développée indépendamment par (Hotelling, 1933). Comme beaucoup de méthodes multivariées, elle n'a pas été largement utilisée jusqu'à l'avènement des ordinateurs électroniques.

Les cercles de corrélation dans le plan des axes F1 et F2 qui représentent une Ecart-type cumulée de 56.71 (%) (F1 = 30.10%, F2 = 26.61 %) (Fig. 58) montrent :

Une association constituée de (PORE CORE) et le (PHIE), Ces paramètres sont corrélés négativement avec le facteur F1 et le facteur F2.

Une association constituée des résistivités (pole électrique) : AHO10, AHO90, Ces paramètres sont corrélés négativement avec le facteur F2 et le facteur F1.

La perméabilité (PERM CORE) ne corrèle pas avec aucuns paramètres avec le facteur F1 et le facteur F2.

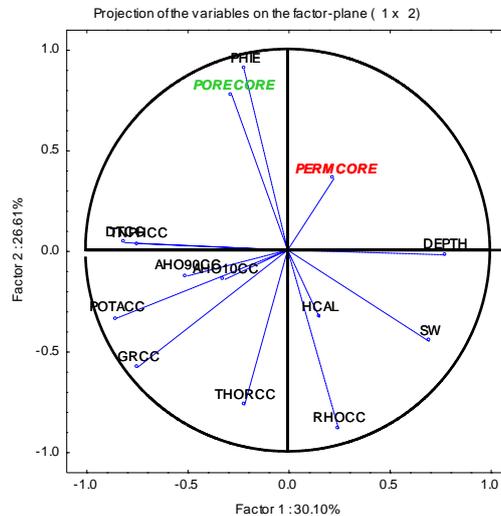


Fig. 58 - Cercle de corrélations de l'ACP sur l'ensemble de données.

L'application de l'ACP a pour objet de rechercher une relation linéaire entre les paramètres à modéliser (la porosité et la perméabilité) et les enregistrements de diagraphies. Les résultats trouvés d'après l'ACP ont montré l'inexistante de relation linéaire entre (PORE CORE, PERM CORE) et les enregistrements de diagraphies. Ce qui nécessite à rechercher une relation non linéaire entre ces paramètres (Zerrouki, 2015), en appliquant les différentes techniques de l'intelligence artificielle (Support Vector Machine, Random Forest, et les réseaux de neurones artificiels).

Dans le présent travail la méthodologie proposée consiste à estimer la porosité et la perméabilité dans les intervalles non carottés à partir de ces paramètres, parce que les enregistrements diagraphiques sont disponibles dans la majorité des puits pétroliers (Zerrouki, 2015).

Les résultats obtenus par ACP montrent le non linéarité entre la perméabilité et la porosité avec les enregistrements de diagraphies dans tout le réservoir des puits étudiés.

1.1.4. Normalisation des données

Lors du traitement des données réelles, en raison des différentes dimensions du paramètre d'évaluation des roches réservoirs, le niveau de données réelles de volume varie considérablement. Si l'on calcule en utilisant directement les données brutes, les données ayant un volume plus important seraient de plus en plus remarquables, alors que l'indicateur avec un volume plus faible et une plus grande sensibilité, sera sous estimée. Ainsi, nous devrions prétraiter et normaliser les données brutes. Dans ce travail, la normalisation des données a lieu en utilisant les valeurs maximales et minimales des données d'après (Vaferi et al., 2011), pour améliorer la convergence du réseau de neurones.

$$X_i = \frac{(X_i - X_{\min})}{X_{\max} - X_{\min}} \quad (14)$$

Telle que X_i est la valeur originale du paramètre, X_i' est la valeur normalisée de X_i , X_{\min} et X_{\max} sont les valeurs minimale et maximale de X_i , respectivement.

1.1.5. Séparation de la base de données

a. Apprentissage

La base de données est scindée en trois parties, la première servant comme base d'apprentissage), et la seconde utilisée comme base de validation (performances du modèle) et la troisième pour (test). Il n'y a pas de règle pour déterminer ce partage de manière quantitatif. Il résulte souvent d'un compromis tenant compte du nombre de données dont on dispose et du temps imparti pour effectuer l'apprentissage. Chaque base doit cependant satisfaire aux contraintes de représentativité de chaque classe de données, et doit généralement refléter la distribution réelle, c'est à dire la probabilité d'occurrence des diverses classes.

b. Validation

Une fois le réseau de neurones entraîné, il est nécessaire de le tester sur une base de données différente de celle utilisée pour l'apprentissage. Ce test permet d'apprécier les performances du modèle élaboré. Si ces performances ne sont pas satisfaisantes, il faudra soit modifier l'architecture du réseau, soit modifier la base d'apprentissage. La démarche expérimentale exige cependant une étape de validation avant toute utilisation. Elle est en effet nécessaire, pour vérifier la fiabilité de la représentation interne du réseau. Elle

consiste en fait à simuler l'utilisation du réseau.

Cette validation est effectuée à l'aide de couples entrées-sorties, autres que ceux utilisés lors de l'apprentissage. Évidemment, le réseau étant par définition un outil d'interpolation, les valeurs des entrées et des sorties de ces couples, tout comme celles des entrées fournies lors de l'utilisation, appartiennent aux mêmes intervalles que celles des expériences qui ont servies à l'apprentissage.

La validation consiste alors à fournir au réseau de neurones les entrées de ces couples, à partir desquelles il calcule (grâce aux poids définis à l'apprentissage) les sorties correspondantes. Ce qui permet de constater l'erreur commise par le réseau de neurones. Cette étape de validation est particulièrement utilisée lors de l'ajustement d'un réseau de neurones à la situation à modéliser, puisqu'il quantifie l'erreur commise. Elle est donc systématiquement réalisée à chaque test effectué lors de la détermination de l'architecture optimale et des autres paramètres (fonction d'activation par exemple); Il s'agit d'une évaluation de l'erreur de généralisation. La différence entre l'erreur d'apprentissage et l'erreur de généralisation représente une mesure de la qualité de l'apprentissage effectué.

c. Test

En réalité il n'y a pas grande différence entre la validation et le test puisque il réalise la même opération, Seulement la phase de validation est utilisée aussi pour éviter le problème de sur apprentissage (Overfitting) et augmenter la capacité de généralisation du modèle. Pendant l'arrêt de l'apprentissage du modèle, on constate que l'erreur totale sur la partie validation cesse de se dégrader, donc à ce point-là l'apprentissage est stoppé ce qui s'appelle l'Early Stopping en Anglais.

2. Prédiction de la Perméabilité par des Différentes Techniques de l'Intelligence Artificielle

La perméabilité est l'un des paramètres les plus importants et les plus difficiles dans la caractérisation des réservoirs pétroliers, car elle ne peut pas être directement mesurée par la technologie actuelle de diagraphie, malgré cela, elle doit être estimée avec précision pour une production optimale de pétrole et du gaz. La perméabilité peut être obtenue à partir une méthode directe au laboratoire sur des échantillons de carottes par l'injection d'air sec ou par les méthodes des essais de puits. Ces méthodes ont un coût très élevé et prennent beaucoup de temps.

En raison de l'importance de ce paramètre de nombreux chercheurs ont tenté de déterminer la perméabilité, par l'établissement des relations avec plusieurs propriétés mesurables de la roche (McGowen & Bloch, 1985; Atkinson et al., 1990). La géométrie des pores, la taille des grains, la minéralogie et la saturation irréductible en eau (Timur, 1968; Walsh & Brace, 1984), L'influence de la composition de la roche sur la perméabilité est également signalée par plusieurs chercheurs (Seeman & Scherer, 1984). D'autres paramètres, utilisés pour estimer la perméabilité, comprennent la surface de l'espace des pores (Carman, 1937; Schwartz & Banavar, 1989), facteur formation et la résistivité (Archie, 1942), pression capillaire (Serra, 1984; Rezaee et Lemon, 1997).

Ces approches avaient des limites et ne sont pas universellement applicables, malgré leur mise en œuvre simple. Une nouvelle tendance a été suscitée dans l'application de systèmes intelligents dans le domaine pétrolier qui témoignent d'une efficacité cohérente (Asoodeh & Bagheripour, 2012; Zargar et al., 2014). Un développement récent vise à prédire perméabilité à partir de données de diagraphie de puits en utilisant des techniques de réseau de neurones (Mohaghegh et al., 1995; Huang et al., 1996; Kadkhodaei-Ilkchi et al., 2005; Zhang et al., 2006).

Dans cette partie de thèse, on propose trois méthodes de l'intelligence artificielle qui sont (Support Vector machine, Randon Forest et les réseaux de neurones artificiels) pour la prédiction de la perméabilité dans le réservoir hétérogène (TAGI) à partir de 332 points données de diagraphie et de perméabilité mesurés.

2.1. Application du réseau de neurones artificiels

Le modèle de réseau de neurones proposé pour la prédiction de la perméabilité est de type perceptron multicouche, appliqué sur l'ensemble de la base de données normalisée pour toutes les données des 3 puits carottés (A, B, C). La base de données est subdivisée aléatoirement en trois parties, 70% des données sont utilisées pour la phase d'apprentissage et 15% pour la validation et 15% pour test.

Pour l'architecture du réseau de neurone proposé est composé d'une couche d'entrée constituée de 13 variables explicatives qui sont les diagraphies (DEPTH, TNPH, DT, RHOB, AHO10, AHO90, GR, POTA, THO, HCAL, et PHIE, S_w). Une seule couche cachée, constituée de 13 neurones avec une fonction de transfert sigmoïdale. La couche de sortie est constituée d'un seul neurone qui est le variable a expliqué (la perméabilité) (Fig. 59).

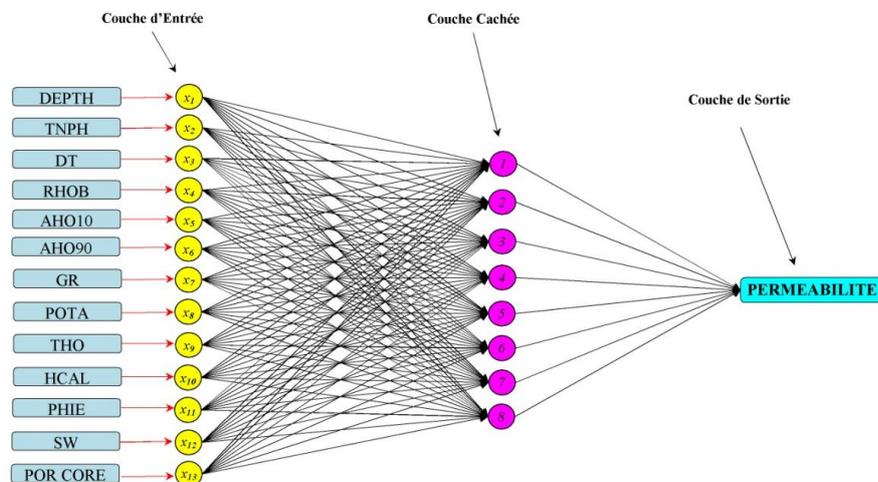


Fig. 59- Architecture de réseau de neurone approprié pour estimer la perméabilité.

L'algorithme utilisé dans l'apprentissage est l'algorithme de rétropropagation de l'erreur, c'est un algorithme qui a donné de bons résultats dans différents domaines. Le réseau de neurones artificiels avec différents nombres de neurones dans la couche cachée est entraîné.

Les poids sont aléatoirement initialisés, on a pris une seule couche cachée et on a fixé le nombre d'itérations tout en changeant le nombre de neurones dans la couche cachée. Le nombre de neurones est choisi par la méthode trial and error, les résultats des essais d'amélioration des performances des réseaux ont été reporté dans le tableau. 20, on a testé les performances des modèles par les indices numériques (RMSE, MAE, et R^2).

Remarque : On a changé le nombre de neurone dans la couche cachée de 1 à 20 neurones et on a uniquement mis dans quelques changements.

Tableau. 20 : Performances des modèles de réseaux de neurones élaborés.

Nombre de neurones	R ²	RMSE	MAE
2	0.84	0.0975	0.0697
5	0.90	0.0779	0.0560
8	0.92	0.0613	0.0348
13	0.85	0.0956	0.0650
17	0.87	0.0884	0.0651
19	0.88	0.0886	0.0608

D'après le Tableau 20, le réseau de neurones artificiels (RNA) montre sa capacité d'apprendre dans la phase d'apprentissage, la meilleure performance est trouvée avec une structure de 8 neurones dans la couche cachée et un nombre d'itérations égal à 200. Le coefficient de corrélation, l'erreur quadratique moyenne et l'erreur absolue sont égaux à 0.92 0.0613 et 0.0348 respectivement.

On remarque que la perméabilité mesurée expérimentalement sur les carottes au laboratoire et la perméabilité calculé par RNA avec la topologie MLP (13-8-1) (Fi. 59). Coïncident avec la perméabilité au laboratoire (Figure. 60). Cette méthodologie peut être appliquée dans d'autres réservoirs gréseux et des puits voisins.

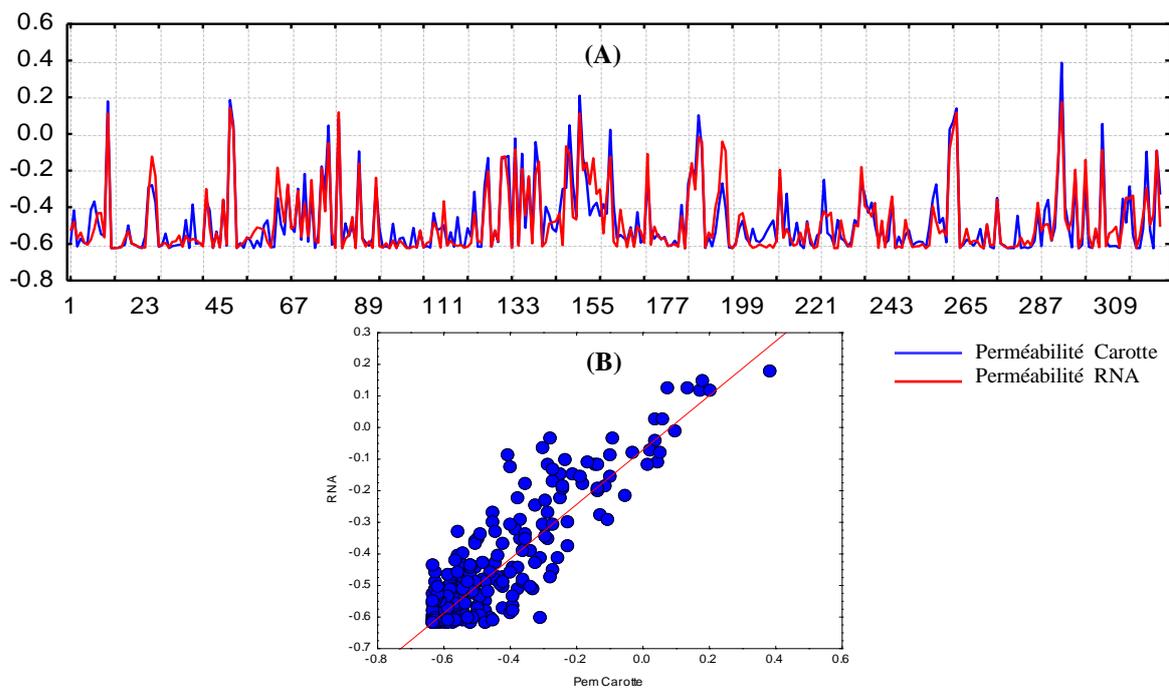


Fig. 60 - Perméabilité carotte (réelle) par rapport à la perméabilité estimée par RNA.

Pour tester et comparer l'efficacité des réseaux de neurones dans la prédiction de la perméabilité on a choisi deux autres techniques de l'intelligence artificielle qui sont largement appliquées pour estimer la perméabilité qui sont support vector machine et random forest.

Les résultats suivants montrent les performances pour chaque technique.

2.2. Application de Support Vector Machine (SVM)

On a proposé la méthode de SVM pour estimer la perméabilité dans le réservoir de TAGI, comme une autre technique de l'intelligence artificielle qui repose sur un algorithme d'apprentissage supervisé basé sur la théorie d'apprentissage statistique, grâce à sa large utilisation dans l'estimation de la perméabilité avec succès surtout dans les réservoirs carbonatés moyennement hétérogènes (Ahmadi, 2018) avec les performances, $R^2 = 0.99$, $RMSE = 0.4176$.

Pour l'élaboration du modèle prédictif de perméabilité par la méthode SVM, on a utilisé l'ensemble de la base de données composée de 13 variables, avec la technique d'évaluation N Cross Validation (Leave One Out), pour le choix des paramètres du noyau nous nous sommes basés sur des tentatives d'amélioration (Trial and Error) à chaque fois on change les paramètres de noyau et de vérifier les performances numériques entre la perméabilité carotte avec simulé et prendre le noyau le plus approprié pour l'estimation de la perméabilité.

Les meilleures paramètres du modèle élaboré est de type 1 SVM (capacity=10.000, epsilon=0.100) Noyau : Fonction de base radiale (gamma=0.600) Nombre des support vectors = 131 (49 bounded), l'erreur de la Cross-validation error est de 0.053, l'erreur quadratique moyenne (RMSE) et le coefficient de détermination (R^2) de la phase d'apprentissage et test sont ; (RMSE 0.006 et 0.01), (CC 0.923, 0.765) =respectivement,

Le coefficient de corrélation ajusté pour = 0.923, $RMSE = 0.1251$, et $MAE = 0.0853$. La figure (61) montre une légère concordance entre la perméabilité réel avec la perméabilité SVM, dans notre cas elle du réservoir hétérogène elle ne donne pas une grande performance.

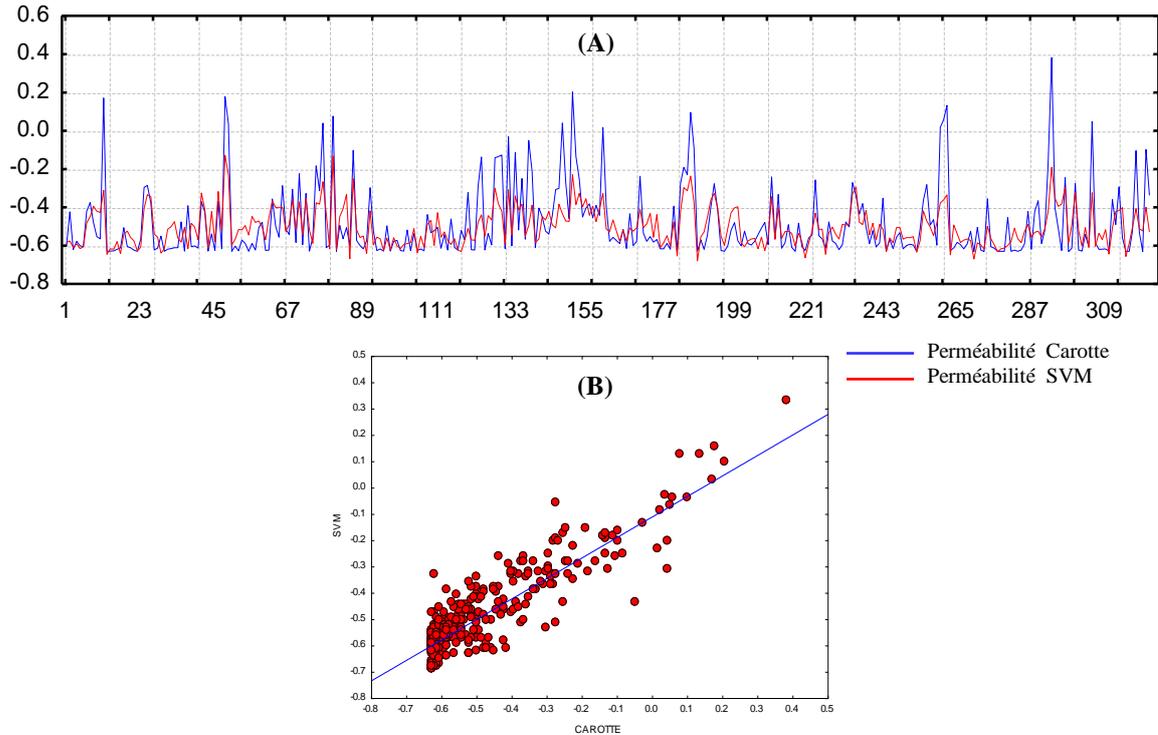


Fig. 61 - Perméabilité carotte (réelle) par rapport à la perméabilité estimée par SVM.

2.3. Application de la méthode des Forêts Aléatoires (Random Forest)

Les forêts aléatoires sont des méthodes qui permettent d'obtenir des modèles prédictifs pour la classification et la régression. La méthode met en œuvre des arbres de décision binaire, notamment des arbres CART proposés par [Breiman et al. \(1984\)](#).

On va proposer cette technique pour réaliser une régression sur notre base de données décrite par des variables qualitatives qui sont les diagraphies de puits et l'objectif (la perméabilité) pour tester si elle donne des bons résultats. L'idée générale de cette technique au lieu d'essayer d'obtenir une méthode optimisée en une fois, on génère plusieurs prédicteurs avant de mettre en commun leurs différentes prédictions, en régression (variable réponse continue) : la méthode permet de prédire la valeur prise par une variable quantitative dépendante, en fonction de variables explicatives.

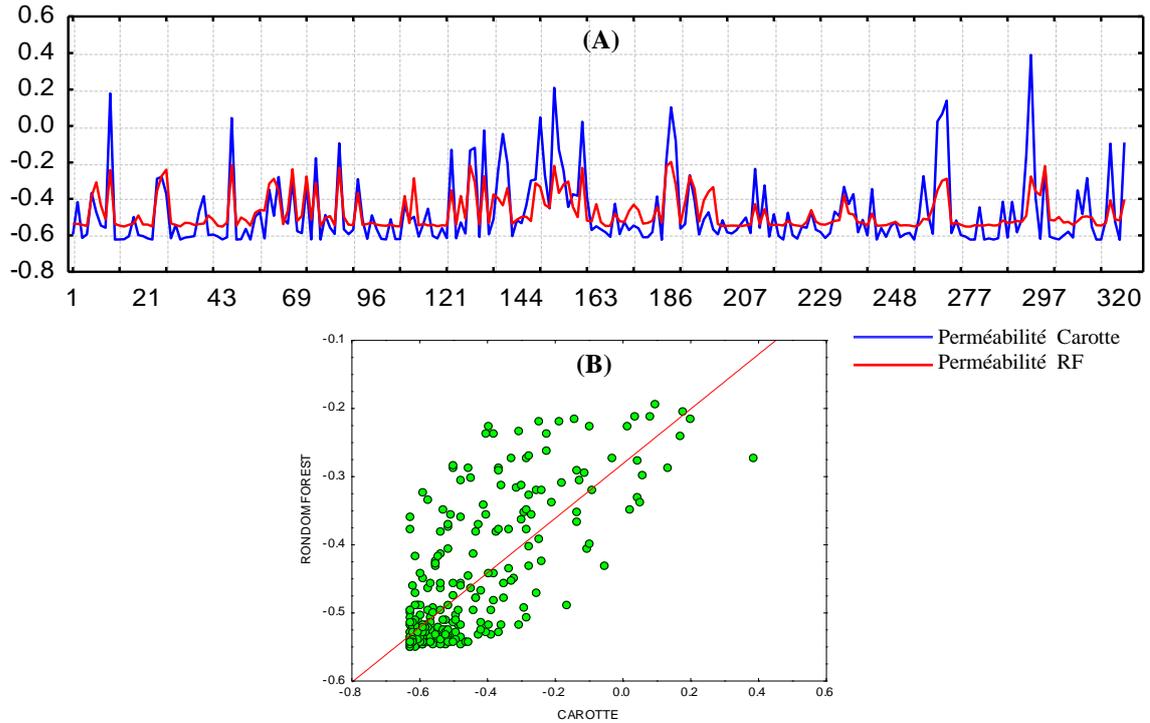


Fig. 62 - Perméabilité carotte (réelle) par rapport à la perméabilité estimée Randon Forest.

Les performances de cette méthode sont faibles R^2 0.790, RMSE 0.118 MAE 0.082, par rapport les deux autres méthodes.

D'après la Figure. 62, il y a une dispersion entre la perméabilité carotte et la perméabilité calculé (RF), ce qui prouve une faible efficacité de cette méthode dans la prédiction de la perméabilité.

2.4. Comparaison entre la perméabilité carotte et la perméabilité simulée par les différentes techniques de l'intelligence artificielle

Cette section présente une comparaison numérique et graphique entre la perméabilité mesurée au laboratoire avec les résultats simulés des trois approches appliquées qui sont : les réseaux de neurones, Support Vector machine et Randon Forest. Le tableau (21) résume les performances de chaque méthode.

Tableau. 21 : Critères de performance de la perméabilité pour chaque méthode.

Méthodes	R^2	RMSE	MAE
Réseaux de Neurones	0.920	0.0613	0.0348
Support Vector machine	0.776	0.1251	0.0853
Random Forest	0.790	0.1180	0.0820

D'après la figure ci-dessous des box-plots et les régressions, il apparaît clairement que les résultats des réseaux de neurones et la plus fiables que les autres techniques.

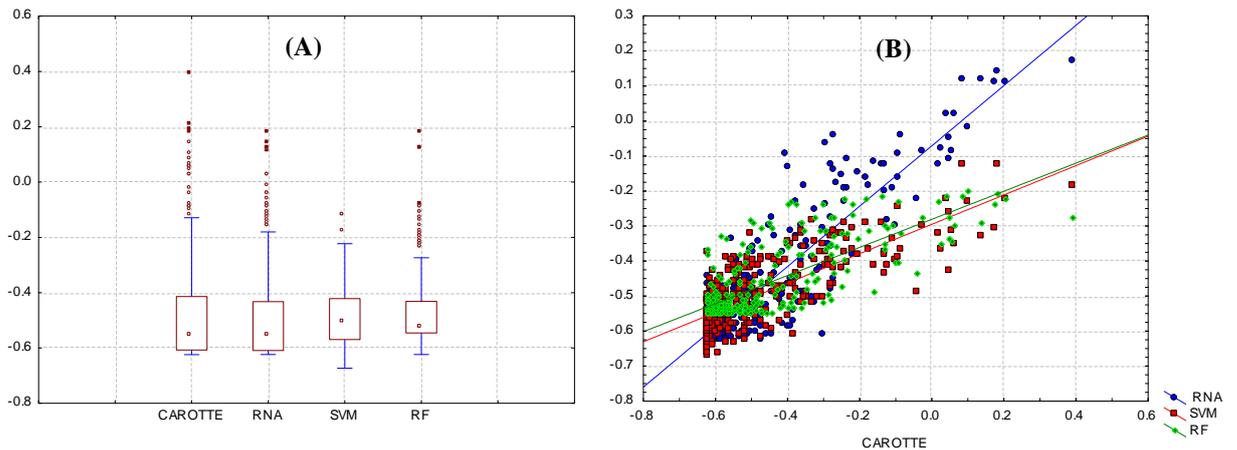


Fig. 63 - Comparaison entre la perméabilité carotte et la perméabilité simulée par les différentes techniques de l'intelligence artificielle.

D'après les résultats numériques et graphiques (Fig. 63) la technique des réseaux de neurones a montré l'efficacité pour la prédiction de la perméabilité dans réservoir de TAGI dans le champ de Sif Fatima.

3. Prédiction de la Porosité par des Différentes Techniques de l'Intelligence Artificielle

La porosité est considérée comme l'un des paramètres les plus importants dans l'estimation des réserves de pétrole et de gaz, car elle détermine le volume effectif des hydrocarbures stockés dans le réservoir. Certainement sur la base des réserves en place, le plan de développement et la gestion du gisement peuvent être définis.

La porosité peut être déterminée par les méthodes directes au laboratoire mais elle reste dépendante de la disponibilité des carottes, le temps et très coûteuse. Cependant, elle peut également être estimée indirectement à partir des diagraphies classiques telles que la densité, neutron, sonique et résonance magnétique, malgré cela il y a beaucoup d'incertitude dans l'estimation de la porosité à partir de diagraphies ce qui implique sur les réserves en place, soit par une sous-estimation ou surestimation qui provoque des grandes problèmes dans l'investissement et le développement du gisement.

Alors la prédiction de la porosité à partir les diagraphies en utilisant les différentes approches de l'intelligence artificielle (IA) et validées avec les valeurs mesurées au laboratoire est le meilleur moyen pour déterminer une valeur précise de la porosité dans tout le réservoir, même dans les puits qui sont totalement non carottés.

L'objectif de notre étude est d'évaluer des outils d'IA tels que l'algorithme génétique combinée avec la régression linéaire multiple (GA-RLM), le réseau de neurones artificiels (RNA), le support machine vectorielle (SVM) Radome Forest (RF) pour prédire le réservoir porosité basée sur les données diagraphies et de 332 mesures de porosité réelle au laboratoire.

3.1. Application du modèle Hybride d'Algorithmes Génétiques et Régression Linéaire Multiple (AG-RLM)

Les algorithmes génétiques sont des algorithmes d'optimisation s'appuyant sur des techniques dérivées de la génétique et de l'évolution naturelle: croisements, mutations, sélection, etc. Les algorithmes génétiques ont déjà une histoire relativement ancienne des premiers travaux de ([John Holland 1992](#)) sur les systèmes adaptatifs. (*Voir chapitre I*).

Cette technique a une capacité de la sélection qui permet d'identifier statistiquement les meilleurs individus d'une population et d'éliminer les mauvais ([Broadhurst, 1997](#)). On trouve dans la littérature un nombre important de principes de sélection plus ou moins adaptés aux problèmes qu'ils traitent. Dans le cadre de notre travail, l'approche hybride algorithmes génétiques avec la régression linéaire multiple (AG-RLM) a été appliquée sur l'ensemble de 13 paramètres de la base de données pour la prédiction et la sélection des variables explicatives les plus influencés sur la prédiction de la porosité dans le réservoir de TAGI. L'équation de (AG-RLM) a été reportée sur la relation suivante.

$$\text{Porosité} = - 0.3110 (\pm 0.1001) \text{HCAL} - 25.8173 (\pm 3.0185) \text{RHOB} + 32.3233 (\pm 7.2865) \text{TNP} + 77.5982 (\pm 7.7970).$$

Les performances de cette technique entre la porosité réelle et la porosité estimée par (AG-RLM) sont $R^2= 0.90$, $\text{RMSE}= 1.2438$ et $\text{MAE}= 1.0308$.

La validation croisée par « leave-one-out » ([Allen, 1974](#)), consiste à recalculer le modèle sur (n-1) objets, et à utiliser le modèle ainsi obtenu pour prédire la valeur de la variable dépendante, le modèle est testé sa capacité prédictive par le critère de validation externe $Q^2=0.82$., et p value =0.0000. Ce qui prouve la capacité prédictive et l'interprétabilité de cette méthode.

Tableau.22 : Paramètres du modèle de (AG-RLM)

Prédicteur	Coefficient	Ecart-type	Confiance 95%	P-value
Constante	77.5982	3.3911	7.7970	0.0000
HCAL	- 0.3110	0.0431	0.1001	0.0000
RHOB	- 25.8173	1.3108	3.0185	0.0000
TNP	32.3233	3.2010	7.2865	0.0000

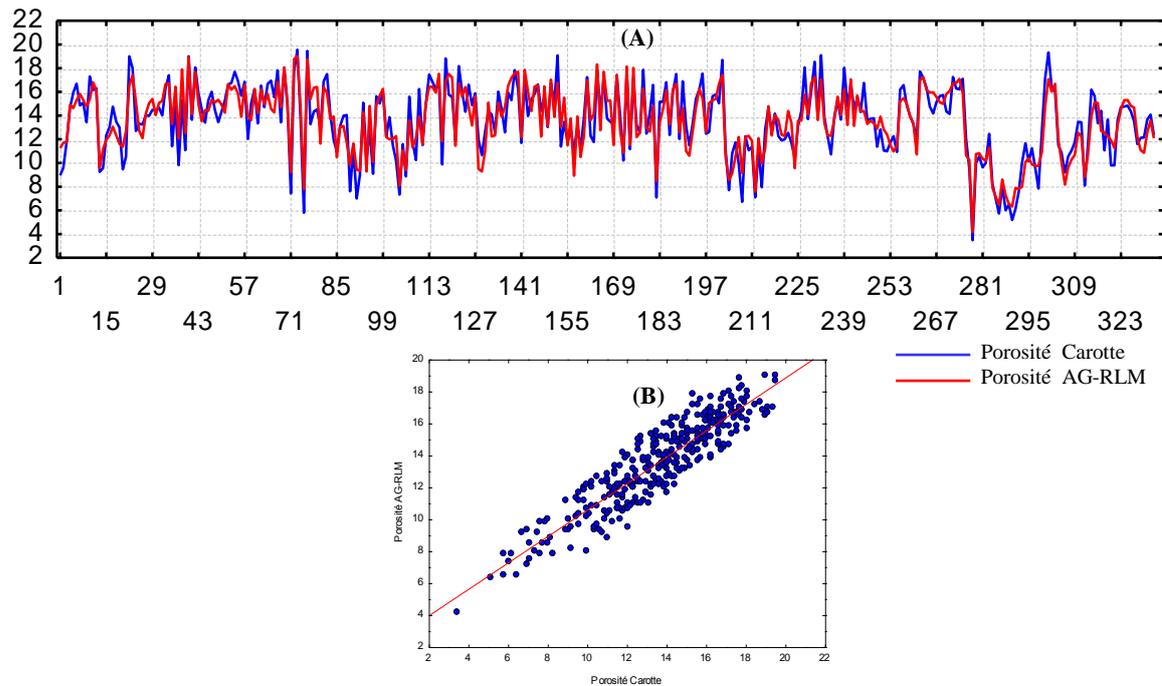


Fig. 64 - Porosité carotte (réelle) par rapport à la porosité estimée par AG-RLM.

Les indices numériques (Tableau. 22) et la visualisation graphique (Fig. 64), ont montré la capacité prédictive de la combinaison de l'algorithme génétique et la régression linéaire multiple pour estimer la porosité dans le réservoir de TAGI.

La sélection des variables

Cette méthode non seulement prédictive mais aussi sert à la sélection des variables les plus significatifs pour estimer la porosité, les variables sélectionnés par (AG-RLM) sont : le Caliper (HCAL), la densité (RHOB) et le neutron (TNPH), ces paramètres qu'il faut les prendre en considération comme des variables explicatifs et les appliquées avec d'autres techniques de l'intelligence artificielle, afin d'améliorer les performances et minimiser le maximum l'erreur.

Le plus important dans cette sélection est le Caliper (HCAL) qui a considéré par (AG-RLM) comme un variable très explicatif avec une grande contribution et sensibilité au modèle ($- 0.3110 (\pm 0.1001)$ HCAL) et à un faible intervalle de d'erreur de (0.1001) par rapport aux RHOB et TNPH, 3.0185, 7.2865 respectivement.

D'après une recherche bibliographique approfondie, nous n'avons pas trouvé ce paramètre (Caliper) pris comme un variable explicatif dans la prédiction de la porosité par les méthodes de l'intelligence artificielle, les paramètres trouvées les plus utilisés sont les diagraphies classiques (la résistivité, sonique, densité, neutron, gamma ray, et la

saturation en eau) (Ja'fari & Moghadam 2012 ; Singh et al., 2015, Baouche & Baddari 2017 ; Ahmadi, 2018). Alors on doit tester l'efficacité de la sélection (AG-RLM) avec les paramètres des diagraphies classiques par autres méthodes de l'intelligence artificielle (Réseaux de Neurones Artificiels, Support Vector Machine, Randon Forest).

3.2. Application des réseaux de neurones artificiels

Le modèle des réseaux de neurones proposé pour la prédiction de la porosité est basé sur la sélection des variables par la méthode (AG-RLM) qui sont le Caliper (HCAL), la densité (RHOB) et le neutron (TNPH) comme des variables explicatives, et on les compare par les modèles qui utilisent les diagraphies classiques comme input (Ja'fari & Moghadam, 2012; Singh et al., 2015 ; Baouche & baddari, 2017 ; Ahmadi, 2019).

Pour la base de données est subdivisée aléatoirement en trois parties, 70% des données sont utilisées pour la phase d'apprentissage et 15% pour la validation et 15% pour test. L'architecture du réseau de neurone proposé est composé d'une couche d'entrée constituée de 3 variables explicatives qui sont les diagraphies (HCAL, TNPH, RHOB) et une seule couche cachée, constituée de 4 neurones avec une fonction de transfère sigmoïdale. La couche de sortie est constituée d'un seul neurone qui est le variable a expliqué (la porosité) (Fig. 65).

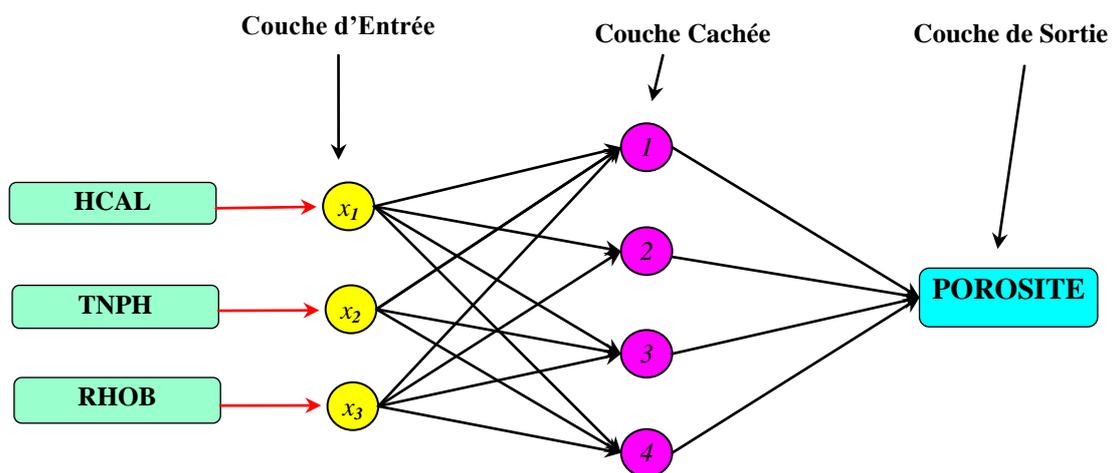


Fig. 65 - Architecture de réseau de neurone approprié pour la prédiction de la porosité.

L'algorithme utilisé dans l'apprentissage est l'algorithme de rétropropagation de l'erreur, c'est un algorithme qui a donné de bons résultats dans différents domaines. Les poids sont aléatoirement initialisés, on a pris une seule couche cachée et on a fixé le nombre d'itérations tout en changeant le nombre de neurones dans la couche cachée. Le

nombre de neurones est choisi par la méthode (trial and error), les résultats de comparaison des indices numériques (RMSE, MAE, et R^2) entre les variables de la sélection (AG-RLM) avec les inputs trouvées dans la bibliographie ont été reportés dans le (Tableau.23)

Tableau. 23 : Performances et architectures des modèles testés.

Input (RNA)	Architectures	R^2	RMSE	MAE
DT, NPFI, RHOB Ja'fari & Moghadam, 2012	3-10 -1	0.87	1.515	1.149
DT, RHOB, Résistivité, Singh et al., 2015	3- 4-1	0.86	1.522	1.218
GR, DT, RHOB, RLLD, PHIE,SW Baouche & Baddari, 2017	6-8 -1	0.88	1.510	1.197
DT, NPFI, RHOB, PHIE Ahmadi, 2019	4-9 -1	0.87	1.532	1.171
HCAL, RHOB, TNPHI	3-4-1	0.92	1.264	1.037

D'après le Tableau (23), les réseaux de neurones artificiels (RNA) montre sa capacité d'estimer la porosité, la meilleure performance est donnée par la combinaison (HCAL, RHOB, TNPHI) par rapport aux autres combinaisons avec une architecture de (3-4-1), et indices de performances ; le coefficient de détermination, l'erreur quadratique moyenne et l'erreur absolue sont égaux à 0.92, 1.412 et 1.109 respectivement.

La figure (66) montre une superposition entre la porosité mesurée expérimentalement au laboratoire avec la porosité calculé par RNA, alors cette méthodologie peut être appliquée dans d'autres réservoirs argileux gréseux ou des puits voisins, pour minimiser les dépenses et avoir des résultats plus précise avec la carotte de $R^2=0.92$.

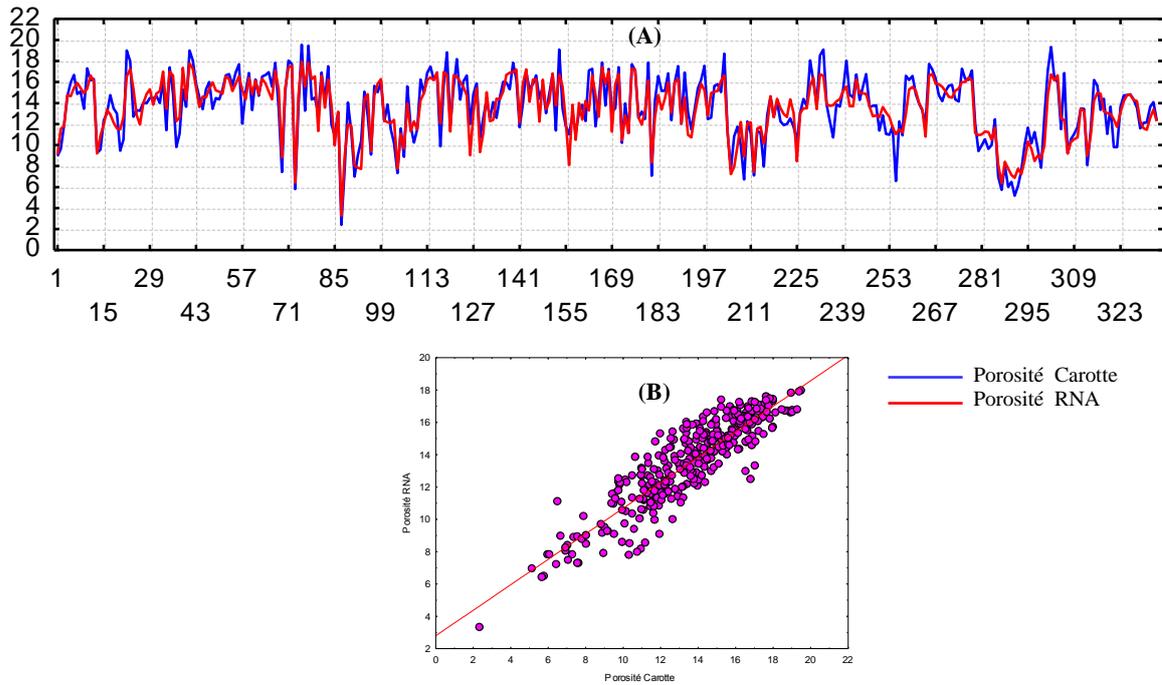


Fig. 66 - Porosité carotte (réelle) par rapport à la porosité estimée par RNA.

Interprétation du Caliper

Le caliper fournit un profil continu du diamètre du puits. Les applications comprennent la localisation d'effondrement ou d'obstruction dans le forage, ainsi que les calculs de volume de ciment, l'identification de la dureté des lithologies présentes le long du puits, la reconnaissance des zones poreuses et perméables (présence de mud-cake), et la Correction des mesures de différentes diagraphies de l'effet du trou ou du mud-cake en vue d'une interprétation plus précise, alors ce paramètre il a une relation étroite avec fonction avec le changement verticale de la porosité.

On recommande ce paramètre vivement de rentrer comme variables explicatifs dans les réservoirs argileux gréseux très hétérogènes pour estimer la porosité par les méthodes de l'IA afin d'avoir une grande précision pour éviter les problèmes de sur ou sous-estimation qui provoque une mal estimation des réserves.

Pour tester et comparer l'efficacité des réseaux de neurones dans la prédiction de la porosité on a proposé deux autres techniques de l'intelligence artificielle qui sont largement appliqués pour estimer la porosité qui sont Support Vector Machine et Random Forest.

3.3. Application de Support Vector Machine (SVM)

Pour l'élaboration du modèle prédictif de porosité par la méthode SVM, on a utilisé les variables sélectionnées par (AG-RLM), avec la technique d'évaluation N Cross Validation (Leave One Out), pour le choix des paramètres du noyau nous nous sommes basés sur des tentatives d'amélioration (Trial and Error) à chaque fois on change les paramètres de noyau et de vérifier les performances numériques entre la porosité carotte avec simulé et prendre le noyau le plus approprié pour l'estimation de la porosité.

Les meilleures paramètres du modèle élaboré sont de type 1 SVM ($C=2.000000$, $\epsilon=0.200000$) Noyau : Fonction de base radiale ($\gamma=0.333333$) Nombre des support vectors = 78 (70 bounded), l'erreur de la Cross-validation error est de 0.00408, Le coefficient de corrélation (R^2) l'erreur quadratique moyenne (RMSE) et l'erreur absolue moyen sont $R^2=0.82$, $RMSE=1.2577$, et $MAE=1.0399$.

La figure (67) montre une légère concordance entre la porosité réelle avec la porosité SVM, dans notre cas elle du réservoir hétérogène elle ne donne pas une grande performance.

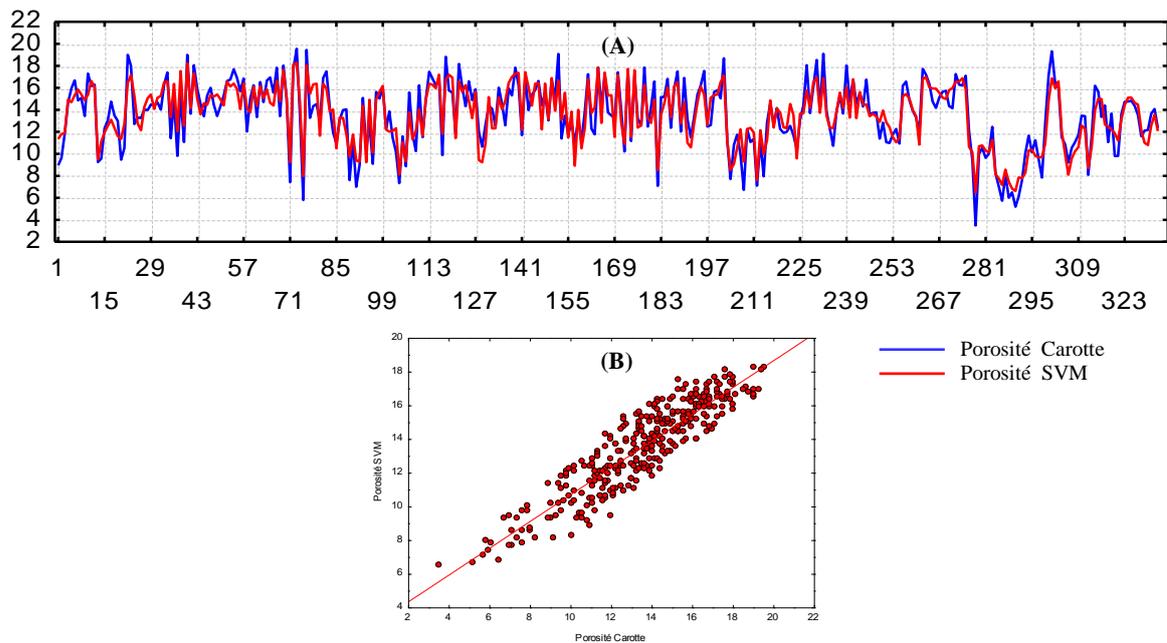


Fig. 67 - Porosité carotte (réelle) par rapport à la porosité estimée par SVM.

3.4. Application de la méthode des Forêts Aléatoires (Rondom Forest)

On va proposer cette technique pour réaliser une régression sur notre base de données décrite par des variables qualitatives qui sont les diagraphies de puits et l'objectif (la porosité) pour tester si elle donne des bons résultats. L'idée générale de cette technique au lieu d'essayer d'obtenir une méthode optimisée en une fois, on génère plusieurs prédicteurs avant de mettre en commun leurs différentes prédictions, en régression (variable réponse continue) : la méthode permet de prédire la valeur prise par une variable quantitative dépendante, en fonction de variables explicatives.

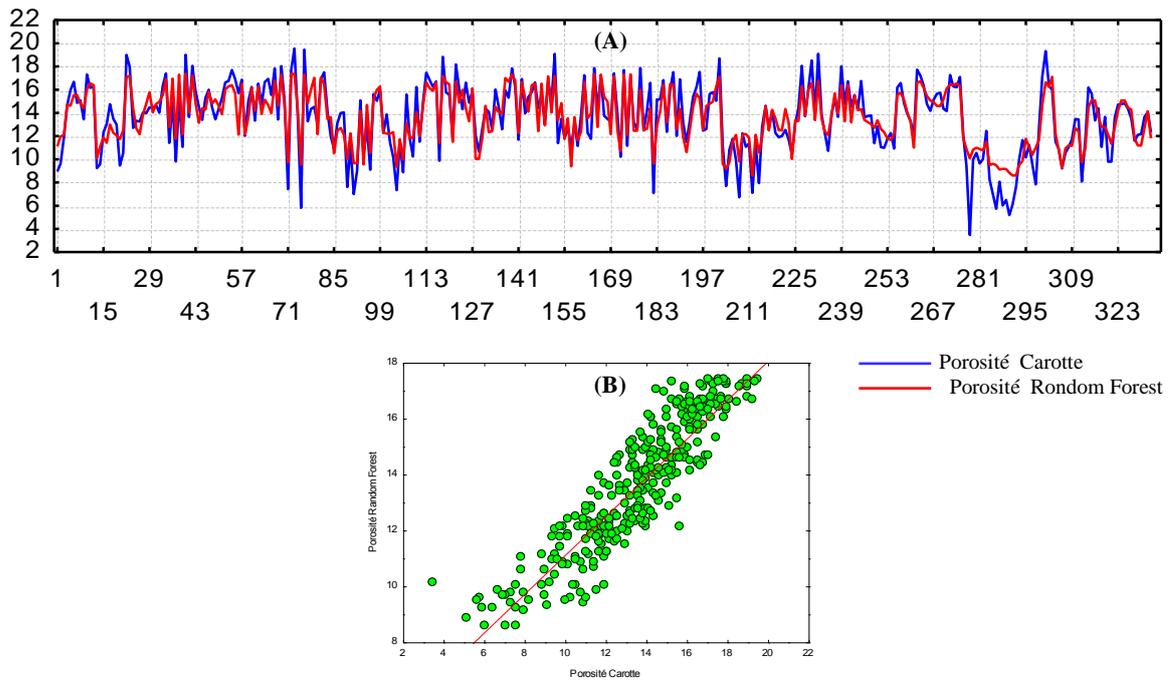


Fig. 68 - Porosité carotte (réelle) par rapport à la porosité estimée par Rondom Forest.

Les performances de cette méthode sont faibles coefficient de détermination $R^2=0.81$, $RMSE=1.353$ $MAE=1.071$, par rapport les deux autres méthodes. Il y a une dispersion entre la porosité carotte et la porosité calculé (RF) (Fig. 68), ce qui prouve une faible efficacité de cette méthode dans la prédiction de la porosité.

3.5. Comparaison entre la porosité carotte et la porosité simulée par les différentes techniques de l'intelligence artificielle

Cette section présente une comparaison numérique par des indices de performances (R^2 , RMSE, MAE) et graphique (les boites à moustaches, régression) entre la porosité mesurée au laboratoire avec les résultats simulés des différentes approches appliquées qui sont : Hybride d'Algorithmes Génétiques et Régression Linéaire Multiple, les réseaux de neurones, Support Vector machine et Randon Forest. Le (tableau 24) résume les performances de chaque méthode.

Tableau. 24 : Critères de performance de la porosité pour chaque méthode

Méthodes	R^2	RMSE	MAE
AG-RLM	0.90	1.2438	1.0308
Réseaux de neurones	0.92	1.2640	1.0370
Support Vector machine	0.82	1.2577	1.0399
Randon Forest	0.81	1.353	1.071

D'après la figure (69) des boites à moustaches et les régressions de la porosité réelle avec la porosité simulée par les différentes techniques, il apparait clairement que les résultats (AG-RLM) et des réseaux de neurones sont plus proches avec la porosité carotte par rapport les deux les autres techniques.

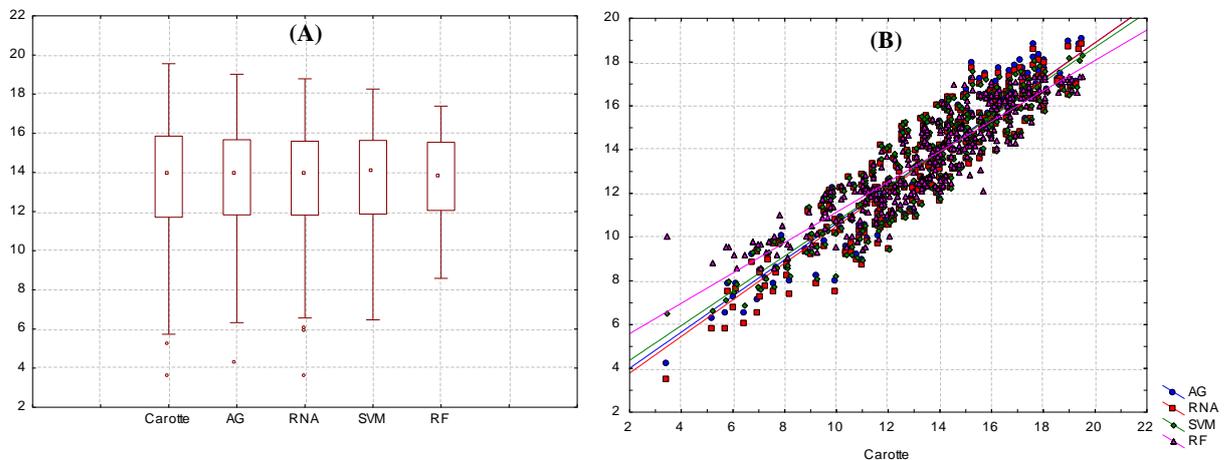


Fig. 69 - Comparaison entre la porosité carotte et la porosité simulée par les différentes techniques de l'intelligence artificielle.

D'après les résultats numériques et la technique hybride d'algorithmes génétiques et régression linéaire multiple et les réseaux de neurones ont montré l'efficacité pour la prédiction de la porosité dans réservoir de TAGI avec des résultats très proches.

4. Généralisation des Modèles de Porosité et de Perméabilité

Le test d'applicabilité et de généralisation du deux modèles élaborés est l'étape la plus importante et la décisive pour tester ses rendements, et juger son capacité de prédire la porosité et la perméabilité à partir des données nouvelles qui n'ont pas rentrés dans la phase d'apprentissage. Pour tester la méthode de prédiction réseaux de neurones artificiels, cette technique a été calibrée dans un puits carotté (P-6) et bien testé sur un maximum de valeurs diagraphiques et de données de carottes tirés aléatoirement. Un bon ajustement entre les données de carottes et les données de porosité et de perméabilité simulées bien remarquées (figure. 70, 71), ce qui suggère une bonne estimation pour la prédiction de ces deux paramètres.

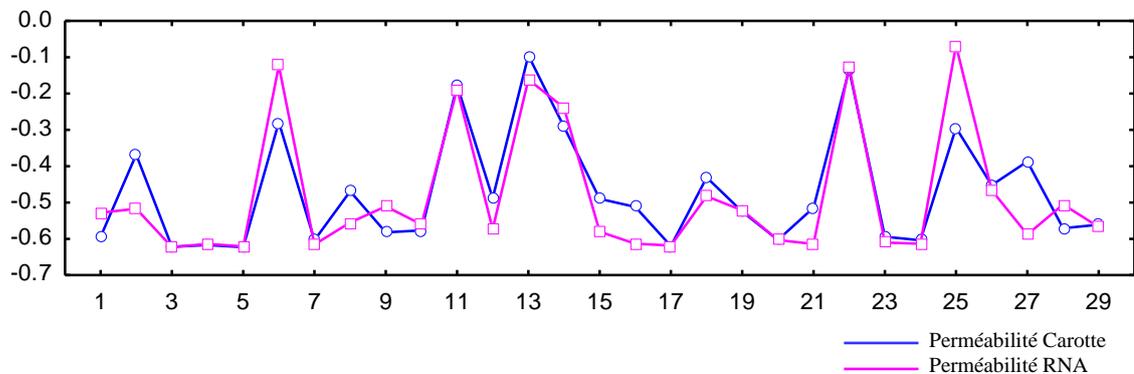


Fig. 70 - Comparaison entre la perméabilité carotte normalisée avec la perméabilité RNA.

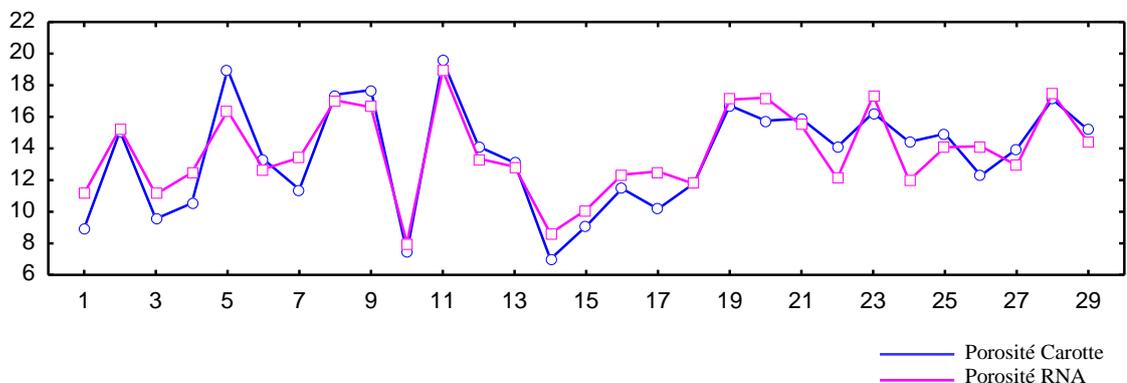


Fig. 71 - Comparaison entre la porosité carotte avec la porosité RNA.

Ils ont prouvé ses capacités marqués par la valeur élevée du coefficient de détermination calculé entre porosité carotte et la porosité RNA et qui est de 0.92, et perméabilité carotte et la perméabilité RNA, $R^2=0.88$.

Conclusion

Quatre techniques différentes de l'intelligence artificielle (AG-RLM, RNA, SVM et RF) ont été évaluées pour estimer la porosité dans le réservoir hétérogène TAGI dans le champ de Sif Fatima, sur la base des diagraphies disponibles et de 332 mesures réelles de porosité qui ont été utilisées pour entraîner et tester les modèles, les conclusions suivantes résumement les résultats obtenus :

1. L'intégration de l'AG-RLM, RNA, SVM et RF pour déterminer la porosité à partir des données de diagraphies. Les résultats montrent que les résultats du RNA a un meilleur coefficient de corrélation de la porosité réelle et la porosité simulée $R^2=0.92$. Mais pour le RMSE la méthode l'AG-RLM proposée est inférieure par rapport les autres méthodes, alors l'AG-RLM peut être recommandée pour prédire la porosité dans tous les autres puits du champ avec une grande fiabilité.
2. Le modèle des réseaux de neurones proposé pour la prédiction de la porosité qui basé sur la sélection des variables par la méthode (AG-RLM) a été comparé par les modèles (Ja'fari & AMoghadam 2012 ; Singh et al., 2015, Baouche & Baddari 2017 ; Ahmadi, 2019) ; le coefficient de corrélation, l'erreur quadratique moyenne et l'erreur absolue sont 0.92, 1.412 et 1.109 respectivement.
3. On recommande vivement le Caliper comme un variable très explicatif dans les réservoirs argileux gréseux très hétérogènes pour estimer la porosité par les méthodes de l'IA afin d'avoir une grande précision pour éviter les problèmes de sur ou sous-estimation qui provoque une mal estimation des réserves.
4. La méthode l'AG-RLM a un gain de temps de calcul, économise du temps et de l'argent dépensé sur le carottage.

Étant donné que la perméabilité est l'un des paramètres les plus importants dans la caractérisation des réservoirs, son estimation dans le l'évaluation du gisement et le scénario de production est très important. Dans notre étude, nous avons utilisé des différentes techniques de l'intelligence artificielle pour estimer la perméabilité dans le réservoir hétérogène TAGI champs de Sif Fatima, la technique des réseaux de neurone MLP proposée avec une topologie de (13.8.1) est plus fiable par rapport aux autres

méthodes par les performances : $R^2=0.92$; RMSE= 0.0613 et MAE= 0.0348 avec des données réelles.

Les quelques erreurs de prédiction des limites d'erreur précises spécifiées se trouvent toujours zones de transition entre lithofacies. Les résultats de cette étude confirment l'efficacité de l'outil des réseaux de neurones dans la prévision de la perméabilité dans le réservoir du trias argileux gréseux du champ de Sif Fatima. Les résultats montrent que la sortie du RNA a les meilleures performances.

CONCLUSION GENERALE

CONCLUSION GENERALE

Le champ pétrolier de Sif Fatima est situé dans l'erg oriental du Sahara algérienne, bassin de Berkine. Cette région est caractérisée par une tectonique cassante relativement intense, elle se manifeste par la présence d'une faille principale d'Ourhoud-Sif Fatima orientée NNE/SSW. Le réservoir essentiel du secteur est celui du Trias Argileux Gréseux Inferieur (TAGI) où les roches réservoirs constituent les séries basales du mésozoïque. Elles sont marquées par les dépôts de nature fluviatile qui se développent le long du bassin. Dans ce réservoir, deux principales séquences peuvent être distinguées à savoir : (1) *le TAGI inférieur* caractérisé par une épaisseur contrôlé par la paléogéographie hercynienne, et (2) *le TAGI supérieur et moyen* où les formations des chenaux et les méandres sont prédominantes. Les séquences terminales *du TAGI supérieur* dans la partie centrale et vers le nord-est du bassin sont marquées par le début d'une transgression marine à la base des calcaires triasiques.

Le réservoir TAGI connu par sa grande hétérogénéité, représente la zone d'étude et l'objectif de la thèse qui focalise sur l'application efficace des techniques de l'intelligence artificielle pour résoudre les problèmes du non récupération des carottes et l'estimation des mesures de la porosité et la perméabilité manquantes. A l'issue de cette thèse, les principaux résultats sont :

I. Prédiction des faciès non carottés :

L'application du MLPNN et les méthodes de *cluster analysis* pour prédire les lithofaciès non carottés dans le réservoir TAGI de cinq puits (trois puits totalement carottés et deux autres totalement non carottés). L'application de ces techniques sur les données des diagraphies nucléaires combinées avec les lithofaciès observés dans les puits carottés. Les résultats obtenus permettent de mettre en évidence que:

1. Les modèles MLPNN et *cluster analysis* doivent être définis sur des données de diagraphies nucléaires en entrée (Input): gamma ray, densité, thorium et potassium qui représentent avec précision un proxy des lithofaciès tandis que d'autres (par exemple, résistivité et sonique) peuvent inclure des interférences en raison de leur sensibilité au fluide interstitiel et à la fracturation de la roche.
2. Le modèle de MLPNN avec quatre entrées, quatre neurones dans la couche cachée et les lithofaciès en sortie représente la meilleure architecture pour prédire les lithofaciès non carottés. Cette architecture (4-4-4) est réalisée avec l'algorithme d'apprentissage

de rétropropagation basé sur Levenberg-Marquardt et la fonction de transfert sigmoïde où le modèle MLPNN atteint les meilleures performances lors de l'apprentissage, de la validation et du test (corrélation R^2 : de 0,78 à 0,92; RMSE: de 0,39 à 0,55; MAE: de 0,23 à 0,30).

3. La comparaison entre le MLPNN et les modèles de *cluster analysis* a prouvé que le MLPNN est le meilleur pour prédire les électrofaciès. La comparaison entre les faciès réels et les électrofaciès simulés indique que le MLPNN reconstitue bien le faciès de grès et de siltstone tandis que des petites interférences dans l'argile, le grès-argileux peuvent se produire. Au contraire, les interférences avec l'emploi du *cluster analysis* sont plus prononcées avec des performances plus faibles ($R^2 = 0,68$, RMSE = 1,04 et MAE = 0,54).
4. L'application de ces méthodes a permis de reconstruire 333 m de faciès non carottés dans le champ pétrolifère de Sif Fatima mettant en évidence une approche intelligente, rapide et totalement économique pour les études souterraines. Ils aident les compagnies pétrolières à réduire les coûts des opérations de carottage.

II. Prédiction de la porosité et la perméabilité:

Pour estimer la porosité, quatre techniques différentes de l'intelligence artificielle (AG-RLM, RNA, SVM et RF) ont été évaluées dans le réservoir TAGI, sur la base des diagraphies disponibles et de 332 mesures réelles de porosité qui ont été utilisées pour entraîner et tester les modèles. Les principaux résultats se résumer en ce qui suit :

1. L'intégration de l'AG-RLM, RNA, SVM et RF a été faite pour déterminer la porosité à partir des données de diagraphies. Les résultats montrent que les résultats du RNA donnent le meilleur coefficient de corrélation entre de la porosité réelle et la porosité simulée ($R^2=0.92$). Cependant, le RMSE de la méthode l'AG-RLM proposée est inférieure par rapport les autres méthodes, alors que l'AG-RLM peut être recommandée pour prédire la porosité dans tous les autres puits du champ avec une grande fiabilité.
2. Le modèle des réseaux de neurones proposé pour la prédiction de la porosité basé sur la sélection des variables par la méthode (AG-RLM) a été comparé par les modèles à partir de la littérature récente (Ja'fari & Moghadam, 2012 ; Singh et al., 2015 ; Baouche & Baddari, 2017 ; Ahmadi, 2019). Dans ce modèle, le coefficient de détermination, l'erreur quadratique moyenne et l'erreur absolue sont 0.92, 1.412 et 1.109 respectivement.

3. Il est recommandé que le Caliper doive être introduit comme une variable très explicative dans les réservoirs argileux gréseux très hétérogènes pour estimer la porosité par les méthodes de l'IA. Ceci va permettre l'obtention d'une grande précision pour éviter les problèmes de surestimation ou sous-estimation qui provoque une mal estimation des réserves.
4. La méthode l'AG-RLM implique le gain du temps de calcul, de l'argent dépensé pour la projection des opérations de carottage si coûteuses.

Pour la perméabilité, étant donné qu'elle est l'une des paramètres les plus importantes lors de la caractérisation des réservoirs. Son estimation pour l'évaluation du gisement et le scénario de production est d'importance primordiale. La technique des réseaux de neurone MLP proposée avec une topologie de (13-8-1) est la plus fiable par rapport aux autres méthodes avec des bonnes performances à savoir : $R^2=0.92$; RMSE= 0.0613 et MAE= 0.0348 en faisant la comparaison avec des données réelles.

Ces méthodes présentent les avantages suivants :

- Le gain économique car ils sont moins coûteuses par rapport l'opération de carottage et les mesures au laboratoire.
- La simulation des paramètres des réservoirs par les techniques de l'intelligence artificielle présent un intérêt rapide quant au gain du temps.

Recommandations et perspectives

A l'issu de cette étude, l'application des différentes techniques de l'intelligence artificielle, pour résoudre des problèmes complexes est très encourageante, et peut encore minimiser les dépenses de carottage et des mesures de laboratoire, une ré-exploitation des data disponibles, par ses méthodes peuvent valoriser d'avantage les connaissances des réservoirs.

Dans le domaine de l'ingénierie des réservoirs, les mathématiques et l'outil informatique sont le locomotif de la modélisation qui ne cessent pas à évoluer d'un jour à l'autre avec des nouvelles approches (combinaison, intégration, hybridation,..). Il faudrait suivre l'actualité des nouveaux techniques et de les appliquer pour résoudre les problèmes de réservoir, pour une meilleure optimisation et évaluation des potentialités en hydrocarbures de notre sous-sol, et même l'état a focalisé sur l'intelligence artificielle de l'appliquer pour une nouvelle ère dans le développement de l'industrie de pétrole en Algérie.

BIBLIOGRAPHIE

BIBLIOGRAPHIE

- Adibifard, M., Tabatabaei-Nejad, S. A. R., Khodapanah, E., 2014. Artificial Neural Network (ANN) to estimate reservoir parameters in Naturally Fractured Reservoirs using well test data. *Journal of Petroleum Science and Engineering* 122, 585-594. <https://doi.org/10.1016/j.petrol.2014.08.007>
- Ahmadi, M.A., Chen, Z., 2018. Development of machine learning methods for estimating permeability and porosity of oil reservoirs via petro-physical logs, *Petroleum*(2018), doi: 10.1016/j.petlm.2018.06.002
- Albus, J. S. 1991. Outline for a theory of intelligence. *IEEE Trans. Systems, Man and Cybernetics*, 21(3):473–509, 1991.
- Ali-Khoudja, S., Chellat, S., Hacini, M., 2020. Petrography and authigenic chlorite in the Siegenian reservoir rocks, Berkine Basin, eastern Algerian Sahara. *Arab J Geosci* **13**, 767 (2020). <https://doi.org/10.1007/s12517-020-05759-z>
- Aliouane, L., Ouadfeul, S.A., Djarfour, N., Boudella, A., 2012. Petrophysical Parameters Estimation from Well-logs Data Using Multilayer Perceptron and Radial Basis Function Neural Networks. In: Huang T., Zeng Z., Li C., Leung C.S. (Eds) *Neural Information Processing. ICONIP 2012. Lecture Notes in Computer Science* 7667. https://doi.org/10.1007/978-3-642-34500-5_86
- Allen, J. G., & Hamsher, J. H., 1974. The development and validation of a test of emotional styles. *Journal of Consulting and Clinical Psychology*, 42(5), 663.
- Al-Mudhafar, W.J., Bondarenko, M., 2015. Integrating K-means clustering analysis and generalized additive model for efficient reservoir characterization. In: *The 77th EAGE Conference & Exhibition Incorporating SPE EUROPEC*, Madrid, Spain, <https://doi.org/10.3997/2214-4609.201413024>
- Amaefule, J.O., Altunbay, M., Taib, D., Kersey, D.G., Keelan, D.K., 1993. Enhanced reservoir description: using core and log data to identify hydraulic (flow) units and predict permeability in uncored intervals/wells, Paper SPE 26436 68th Annual Technical Conference and Exhibition, Houston, Texas, 3–6 October, 1993, 16 pages. <https://doi.org/10.2118/26436-MS>
- Ameur-Zaimeche, O., Zeddouri, A., Kouadria, T., Kechiched, R., Belksier, M S. 2014a. Use of Cluster Analysis method in log's data processing: prediction and rebuilding of lithologic facies. *International Conference on Environmental Science and Geoscience (ESG '14)* Venice, Italy, March 15-17, 2014 98 – 101.
- Ameur-Zaimeche, O., Zeddouri, A., Heddami, S., Kechiched, R. 2014b. Modélisation et reconstitution des facies non carottés à l'aide des méthodes à base des Réseaux de Neurones Artificiels et Cluster Analysis, cas du réservoir Trias Argileux Gréseux inférieur, champ de Sif Fatima, bassin de Berkine, sud de l'Algérie.. 1^{er} Colloque International sur Hydrocarbures, Energies et Environnement (HCEE'2014), Ouargla, Algeria, Novembre 23-24.
- Ameur-Zaimeche, O. 2014c. Modélisation et reconstitution des facies non carottés par les méthodes statistiques multivariées, cas du cas du réservoir Trias Argileux Gréseux inférieur, champ de Sif Fatima, bassin de Berkine, sud de l'Algérie. Magister, Université Kasdi Merbah Ourgla.
- Ameur-Zaimeche, O., Zeddouri, A., Heddami, S., Kechiched, R., Lithofacies prediction in non-cored wells from the Sif Fatima oil field (Berkine basin, southern Algeria): A comparative study of multilayer perceptron neural network and cluster analysis-based approaches, *Journal of African Earth Sciences*, Volume 166, 2020, 103826, ISSN 1464-343X, <https://doi.org/10.1016/j.jafrearsci.2020.103826>.
- Archie, G. E., 1942. The electrical resistivity log as an aid in determining some reservoir characteristics. *Transactions of the AIME*, 146(01), 54-62.
- Asfahani, J., 2014. Statistical Factor Analysis Technique to Interpret Nuclear and Electrical Well Logging Measurements for Basalt Characterization. Case study from Southern Syria. *Applied Radiation and Isotopes* 84, 33-39. <https://doi.org/10.1016/j.apradiso.2013.09.019>
- Askri, H., Belmecheri, A., Benrabah, B., Boudjema, A., Boumendjel, K., Daoudi, M., ... & Ghomari, A. (1995). *Geology of Algeria*. In *Well Evaluation Conference Algeria* (pp. 1-93).
- Asoodeh, M., & Bagheripour, P., 2012. Prediction of compressional, shear, and stoneley wave velocities from conventional well log data using a committee machine with intelligent systems. *Rock mechanics and rock engineering*, 45(1), 45-63.

- Atkinson, Anthony B., 1990. Poverty, statistics, and progress in Europe. Welfare State Programme Discussion Papers (WSP 060). Centre for Analysis of Social Exclusion, The London School of Economics and Political Science, London, UK.
- Azoug, Y., et al. 2007 - Well Evaluation Conference- WEC Algeria 2007 - Sonatrach-Schlumberger. Edited by Schlumberger; Produced by Lynx Consulting, Inc., Houston, TX, USA. 489 pages.
- Bagheripour Parisa, Asoodeh Mojtaba, 2013. Fuzzy ruling between core porosity and petrophysical logs: Subtractive clustering vs. genetic algorithm–pattern search, *Journal of Applied Geophysics*, Volume 99, 2013, Pages 35-41, ISSN 0926 9851, <https://doi.org/10.1016/j.jappgeo.2013.09.014>.
- Ball, G. H. et Hall, D. J., 1967. An Iterative Method of Multivariate Analysis and Pattern Recognition. *Behavior Science*, 153.
- Baouche, R., Sen, S., & Ganguli, S. S., 2020. Pore pressure and in-situ stress magnitudes in the Bhiret Hammou hydrocarbon field, Berkine Basin, Algeria. *Journal of African Earth Sciences*, 171, 103945. <https://doi.org/10.1016/j.jafrearsci.2020.104038>
- Baouche, R., & Beddari, K., 2017. Prediction of permeability and porosity from well log data using the nonparametric regression with multivariate analysis and neural network, Hassi R'Mel Field, Algeria. *Egyptian journal of petroleum*, 26(3), 763-778.
- Bayes, T. Lii. 1763. An essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, frs communicated by mr. price, in a letter to john canton, amfr s. *Philosophical transactions of the Royal Society of London*, (53): 370–418.
- Belhouchet, H.E., Benzagouta, M.S., Dobbi, A., Alquraishi, A., Duplay, J., A., 2020. New Empirical Model for Enhancing Well Log Permeability Prediction, Using Nonlinear regression method: Case study from Hassi-Berkine Oil Field Reservoir – Algeria, *Journal of King Saud University - Engineering Sciences* (2020), doi: <https://doi.org/10.1016/j.jksues.2020.04.008>
- Benzécri J.P. 1973. L'analyse des données. Dunod, Paris.
- Bezdek, J. C., 1981. "Pattern Recognition with Fuzzy Objective Function Algorithms", Plenum Press, New York. 1981
- Binet, A., 1911. Les idées modernes sur les enfants. Flammarion, Paris.
- Bingham, W. V., 1937. Aptitudes and aptitude testing. Harper & Brothers, New York.
- Boote, D. R., Clark-Lowes, D. D., & Traut, M. W., 1998. Palaeozoic petroleum systems of North Africa. *Geological Society, London, Special Publications*, 132(1), 7-68.
- Boudjemaa, A. 1987. Evolution structurale du bassin pétrolier "triasique" du Sahara Nord oriental (Algérie), thèse à l'université de Paris Sud, Centre d'Orsay, 290p.
- Breiman, L., & Ihaka, R., 1984. Nonlinear discriminant analysis via scaling and ACE. Department of Statistics, University of California.
- Breiman, L., 1996. Bagging predictors. *Machine learning*, 24(2), 123-140.
- Breiman, L., 1997. Arcing the edge. Technical Report 486, Statistics Department, University of California at Berkeley.
- Breiman, L., 2001. Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science*, 16(3), 199-231.
- Broadhurst, D., Goodacre, R., Jones, A., Rowland, J. J., & Kell, D. B., 1997. Genetic algorithms as a method for variable selection in multiple linear regression and partial least squares regression, with applications to pyrolysis mass spectrometry. *Analytica Chimica Acta*, 348(1-3), 71-86.
- Brostaux, Y., 2005. Etude du classement par forêts aléatoires d'échantillons perturbés à forte structure d'interaction (Doctoral dissertation, Faculté Universitaire des Sciences agronomiques de Gembloux).
- Byrnes, A.P., 1994. Measurements of independent and dependent variables for petrophysical properties prediction. Reservoir Quality Assessment and Prediction in Clastic Rocks: M.D., Wilson edition, SEPM short course #301994, p. 231-274 & p.293-31.
- Carman, P. C., 1937. Fluid flow through granular beds. *Trans. Inst. Chem. Eng.*, 15, 150-166.
- Celeux, G., Diday, E., Govaert, G., Lechevallier, Y., and Ralambondrainy, H., 1989. Classification automatique des données, environnement statistique et informatique. DUNOD informatique.
- Chelbeb, S., Mouaici, R., Haroun, S., Kracha, N., Ouachouche, F., 2000. Depositional environment, sequence stratigraphy analysis, and diagenesis of the lower Devonian sandstone

- reservoirs, Berkine basin, Algeria. 4ème Séminaire de Géologie Pétrolière-SGP4 – CRD SONATRACH, Boumerdès, 13 au 15 Novembre 2000.
- Chen, T., and Guestrin C., 2016. A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 785–794, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-4232-2. <http://doi.acm.org/10.1145/2939672.2939785>.
- Ching, D. S. Himmelstein, B. K. Beaulieu-Jones, A. A. Kalinin, B. T. Do, G. P. Way, E. Ferrero, P.-M. Agapow, M. Zietz, M. M. Hoffman, W. Xie, G. L. Rosen, B. J. Lengerich, J. Israeli, J. Lanchantin, S. Woloszynek, A. E. Carpenter, A. Shrikumar, J. Xu, E. M. Cofer, C. A. Lavender, S. C. Turaga, A. M. Alexandari, Z. Lu, D. J. Harris, D. DeCaprio, Y. Qi, A. Kundaje, Y. Peng, L. K. Wiley, M. H. S. Segler, S. M. Boca, S. J. Swamidass, A. Huang, A. Gitter, and C. S. Greene. Opportunities and obstacles for deep learning in biology and medicine. *J. R. Soc. Interface*, 15(141), Apr. 2018. ISSN 1742-5689, 1742-5662. doi: 10.1098/rsif.2017.0387. <http://dx.doi.org/10.1098/rsif.2017.0387>
- Corsini, M.M., 2005. Introduction aux réseaux de neurones Université Victor Segalen France.
- Datta-Gupta, A., Xue, G., Lee, S.H., 1999. Nonparametric Transformations for Data Correlation and Integration: From Theory to Practice, in: Reservoir Characterization: Recent Advances, AAPG Datapages, 1999.
- Dieckmann, V., Pajola, M., Caccialanza, G., 2000. New Definition of Silurian Source Rock Distribution in the Berkine Basin (Algeria) and its Importance for Reservoir Infilling. 4ème Séminaire de Géologie Pétrolière-SGP4 – CRD-SONATRACH- Boumerdès, 13 au 15 Novembre 2000.
- Diersen, S., Lee, E.-J., Spears, D., Chen, P., Wang, L., 2011. Classification of Seismic Windows Using Artificial Neural Networks. *Procedia Computer Science* 4, 1572–1581. <https://doi.org/10.1016/j.procs.2011.04.170>
- Djeffal, A., 2012. Utilisation des méthodes Support Vector Machine (SVM) dans l'analyse des bases de données (Doctoral dissertation, Université Mohamed Khider-Biskra).
- Dramsch, J. S., 2019. Machine Learning in 4D Seismic Data Analysis: Deep Neural Networks in Geophysics. Technical University of Denmark.
- Dramsch, J. S., 2020. 70 years of machine learning in geoscience in review. *Advances in Geophysics*. <https://doi.org/10.1016/bs.agph.2020.08.002>
- Dreyfus G., 1998. Les réseaux de neurones »; Cours Mécanique Industrielle et Matériaux, n°51, École Supérieure de Physique et de Chimie Industrielles de la ville de Paris (ESPCI), Septembre 1998.
- Dubois, M.K., Bohling, G.C., Chakrabarti, S., 2007. Comparison of four approaches to a rock facies classification problem. *Computers and Geosciences*, 33(5), pp.599–617. <https://doi.org/10.1016/j.cageo.2006.08.011>
- Dullien, F. A. L., 1992. Porous Media - Fluid Transport and Pore Structure. 2nd edition, Academic, San Diego, Calif., United States
- Dunn, J. C., 1973. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters", *Journal of Cybernetics*, no 3, pp 32-57. 1973.
- Durand, N., 2004. Algorithmes Génétiques et autres méthodes d'optimisation appliqués à la gestion de trafic aérien. Optimisation et contrôle [math.OC]. INPT, 2004. [fftel-01293722f](https://doi.org/10.1016/j.oftel.2004.08.002)
- Echikh, K. 1998. Geology and hydrocarbon occurrences in Ghadames basin, Algeria, Tunisia, Libya. In D. S. MacGregor, R. T. J. Moody & D. D. Clark-Lowes (Eds.), *Petroleum geology of North Africa*. Geological Society of London special publication, 132. 109- 130.
- Galeazzi, S., Point, O., Haddadi, N., Mather, J., Druesne, D., 2010. Regional geology and petroleum systems of the Illizi–Berkine area of the Algerian Saharan Platform: An overview. *Marine and Petroleum Geology* 27 (2010) 143–178.
- Gauthier, F. J., Hedley, R. J., & Mckenna, S., 2003. The Structural & Tectonic Evolution of the Berkine-Ghadames Basin. In *1st EAGE North African/Mediterranean Petroleum & Geosciences Conference & Exhibition* (pp. cp-8). European Association of Geoscientists & Engineers.
- Ghribi, A., 1999. Répartition des roches éruptives dans le bassin de Berkine et commentaires géologiques. Rapport inédit, Sonatrach. 38 p.
- Hamida, Mohamed Assaad., 2015. Notes de cours : Introduction aux méthodes de Contrôle Intelligent, Learning-univ-ouargla.dz

- Handhal, Amna., M. Alaa M. Al-Abadi, Hussein E. Chafeet, Maher J. Ismail, 2020. Prediction of total organic carbon at Rumaila oil field, Southern Iraq using conventional well logs and machine learning algorithms, *Marine and Petroleum Geology*, Volume 116, 2020, 104347, ISSN 0264-8172, <https://doi.org/10.1016/j.marpetgeo.2020.104347>.
- Harris, W. E., & Racine, R. 1979. Globular clusters in galaxies. *Annual Review of Astronomy and Astrophysics*, 17(1), 241-274.
- Hartigans, 1975. clustering algorithms. John Wiley and Sons, Inc., 1975.
- Heddam, S., 2016. Multilayer perceptron neural network-based approach for modeling phycoyanin pigment concentrations: case study from lower Charles River buoy, USA. *Environ. Sci. Pollut. Res.* 23 (2016), 17210. <https://doi.org/10.1007/s11356-016-6905-9>.
- Hinton, G.E. 1989. Deterministic Boltzmann machine learning performs steepest descent in weight-space. *Neural Computation*, 1 :143–150.
- Hornik, K., 1989. Stinchcombe, M., & White, H. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2 (5), pp. 359-366, 1989.
- Hotelling, H. 1933. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(7), 498-520. <https://doi.org/10.1037/h0070888>
- Huang, Z., Shimeld, J., Williamson, M., Katsube, J., 1996. Permeability prediction with artificial neural network modeling in the Venture gas field offshore eastern. Canada *Geophysics* 61, 422-436
- Irani Rasoul, Nasimi Reza., 2011. Evolving neural network using real coded genetic algorithm for permeability estimation of the reservoir, *Expert Systems with Applications*, Volume 38, Issue 8, 2011, Pages 9862-9866, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2011.02.046>.
- Ja'fari, A., & Moghadam, R. H., 2012. Integration of ANFIS, NN and GA to determine core porosity and permeability from conventional well log data. *Journal of Geophysics and Engineering*, 9(5), 473-481.
- Jamialahmadi, M., Javadpour, F., 2000. Relationship of permeability, porosity and depth using an artificial neural network. *J. PetrolSci. Eng.* 26 (1-4), 235-239. [http://dx.doi.org/10.1016/S0920-4105\(00\)00037-1](http://dx.doi.org/10.1016/S0920-4105(00)00037-1).
- Jianhua, H., Andrew, D., La, C., Jiahao, W., Wenlong, D., 2019. Underschultz, Using neural networks and the Markov Chain approach for facies analysis and prediction from well logs in the Precipice Sandstone and Evergreen Formation, Surat Basin, Australia, *Marine and Petroleum Geology* 101,410-427. <https://doi.org/10.1016/j.marpetgeo.2018.12.022>
- John, H., 1992. Holland. genetic algorithms. *Scientific american*, 267(1), 44-50.
- Johnson S. C., 1967. Hierarchical Clustering Schemes" *Psychometrika*, no 2, pp 241-254.
- Kadi, B., Aït-Ouali, R., Zazoun, R.S., 2013. Comparaison de l'histoire thermique des bassins de Timimoun et de Berkine: impact sur l'évolution des argiles et de la matière organique. *Bulletin du Service Géologique National* Vol. 24, n° 3, pp. 1 - 25, 14 fig., 2 tabl., 2013.
- Kadkhodaei-Ilkchi, A., Rezaee, M.R., Moallemi, S.A., 2005. A New Approach for Prediction of Porosity, Permeability and Rock Types Using Soft Computing Techniques, An Example from Southern Persian Gulf, *Petroleum Geoscience Collaboration Conference*, London 30 November-1 December 375
- Kadurin, A., S. Nikolenko, K. Khrabrov, A. Aliper, and A. Zhavoronkov., 2017. An advanced generative adversarial autoencoder model for de novo generation of new molecules with desired molecular properties in silico. *Mol. Pharm.*,14(9):3098–3104, Sept. 2017. ISSN 1543-8384, 1543-8392. <http://dx.doi.org/10.1021/acs.molpharmaceut.7b00346>
- Kannan, T. D. B., Kannan, G. R., Kumar, B. S., Baskar, N., 2014. Application of Artificial Neural Network Modeling for Machining Parameters Optimization in Drilling Operation. *Procedia Materials Science* 5, 2242-2249. <https://doi.org/10.1016/j.mspro.2014.07.433>
- Kaufman, L., Rousseeuw, P.J., 1990. Finding Groups in Data. Wiley, New York, 342pp. <https://trove.nla.gov.au/work/14656687>
- Kaydani Hossein, Mohebbi Ali, Baghaie Ali., 2011. Permeability prediction based on reservoir zonation by a hybrid neural genetic algorithm in one of the Iranian heterogeneous oil reservoirs, *Journal of Petroleum Science and Engineering*, Volume 78, Issue 2, 2011, Pages 497-504, ISSN 0920-4105, <https://doi.org/10.1016/j.petrol.2011.07.017>.

- Kiselev, V.M., Baikalskaya, B., 2011. Use of the Cluster Analysis and Artificial Neural Network Technology for Log Data Interpretation 4, 453-462.
- Kohonen T., 1982. self-organized formation of topologically correct feature maps. *Biological cybernetics* no 43, pp59-69, reprinted in Anderson & Rosenfeld, Eds, *Neurocomputing: foundations of research*, MIT press, Cambridge Ma.
- Kong, Q., D. T. Trugman, Z. E. Ross, M. J. Bianco, B. J. Meade, and P. Gerstoft., 2019. Machine learning in seismology: Turning data into insights. *Seismological Research Letters*, 90(1):3–14, 2019.
- Koudri, M., 2011. Modèle de mélange Gaussien. Application sur image cytologique. Mémoire de master en Informatique, l'Université Abou Bekr Belkaid Tlemcen.
- Ladlani, I., Houichi, L., Djemili, L., Heddami, S., Belouz, K., 2012. Modeling of daily reference evapotranspiration (ET₀) in the north of Algeria using generalized regression neural networks (GRNN) and radial basis function neural networks (RBFNN): a comparative study. *Meteorol Atmos Phys* (2012) 118:163–178.
- Lance, G.N., & Williams, W.T., 1967. A general theory of classificatory sorting strategies : I. Hierarchical systems. *Computer Journal*, no 9, pp 373-380, 1967.
- Lary, D. J., Alavi, A. H., Gandomi, A. H., and A. L. Walker., 2016. Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, 7(1):3–10. ISSN 1674-9871. doi: 10.1016/j.gsf.2015.07.003.
- Lassouani, L., 1996. Influence des épanchements volcaniques sur la sédimentation triasiques dans le bassin de Ghadamès. Rapport inédit, Sonatrach. 48 p
- Lawrence, J., 1994. *Introduction to Neural Networks*. California Scientific Software Press, Nevada City, CA.
- Lebart L., Morineau A. & Piron M., 2000. *statistique exploratoire multidimensionnelle*. Dunod, 3ème édition, paris.
- Legendre, A. M., 1805. *Nouvelles méthodes pour la détermination des orbites des comètes*. F. Didot.
- Legg, S., & Hutter, M., 2007a. A collection of definitions of intelligence. *Frontiers in Artificial Intelligence and applications*, 157, 17.
- Legg, S., & Hutter, M., 2007b. Universal intelligence: A definition of machine intelligence. *Minds and machines*, 17(4), 391-444.
- Li J. and Castagna J., 2004. Support vector machine (SVM) pattern recognition to AVO classification. *Geophys. Res. Lett.*, 31(2):948, Jan. 2004. ISSN 0094-8276. doi: 10.1029/2003GL018299. <http://doi.wiley.com/10.1029/2003GL018299>.
- Lim, J.-S., 2005. Reservoir properties determination using fuzzy logic and neural networks from well data in offshore Korea. *J. Petrol. Sci. Eng.* 49 (3-4), 182-192. <http://dx.doi.org/10.1016/j.petrol.2005.05.005>.
- MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Vol. 1, No. 14, pp. 281-297).
- Marc, P., 2004. Réseaux de Neurones Université de l'AVAL.
- Markov, A. A., 1906. Rasprostranenie zakona bol'shich chisel na velichiny, zavisyaschie drug ot druga. *Izvestiya Fizikomatematicheskogo obschestva pri Kazanskom universitete*, 15(135-156):18.
- Markov, A. A., 1971. Extension of the limit theorems of probability theory to a sum of variables connected in a chain. *Dynamic probabilistic systems*, 1:552–577. Reprint in English of [Markov, 1906].
- McCulloch, W. S & Pitts, W., 1943. A logical calculus of the ideas immanent in Nervous activity. *Bull Math Biophysics*, 5:113–115.
- McGowen, J H, and Bloch, S., 2002. Depositional facies, diagenesis, and reservoir quality of Ivishak sandstone (Sadlerochit Group), Prudhoe Bay field. United States: N. p., 1985. Web.
- McKenna, S., Hedley, R., 2002. The structural evolution of the Berkine-Ghadames Basin (Groupement Berkine Anadarko). Rapport interne – Anadarko Petroleum Corporation.
- Mohaghegh S, Balan B, Ameri S., 1997. Permeability determination from well log data. *SPE Form Eval* 12(03):170–174. doi:10.2118/30978-PA
- Mohaghegh, S.D., 2005. Recent developments in application of artificial intelligence in petroleum engineering. *J. Petrol. Technol.* 57 (4), 86-91.

- Namig, J. Guliyev, Vugar, E. Ismailov, 2017. On the approximation by single hidden layer feedforward neural networks with fixed weights. *Neural Network*. 98 (2018), 296–304. <https://doi.org/10.1016/j.neunet.2017.12.007>
- Newendorp, P. D., 1976. Decision analysis for petroleum exploration.
- Newman, G.H; Martin, J.C., 1977. Equipment and experimental methods for obtaining laboratory compression characteristics of reservoir rocks under various stress and pressure conditions. SPE paper no. 6855, 52nd Annual Technical Conference of SPE, Denver, Colorado, October 9-12, 1977.
- Ouadfeul, S., Zaourar, N., Boudella, A., & Hamoudi, M., 2011. Modeling and classification of lithofacies using the continuous wavelet transform and neural network: A case study from Berkine Basin (Algeria). *Bulletin du Service Géologique d'Algérie*, 22(1), 1-16.
- Panzoli, D., 2008. Proposition de l'architecture "Cortexionist" pour l'intelligence comportementale de créatures artificielles (Doctoral dissertation, Université de Toulouse, Université Toulouse III-Paul Sabatier).
- Parsaye K. A., 1998. Characterization of Data Mining Technologies and Processes / K. A. Parsaye // *The Journal of Data Warehousing* 1, 41-63.
- Pearson, K., 1901. On Lines and Planes of Closest Fit to Systems of Points in Space. *Philosophical Magazine*, 11, 559-572. <http://dx.doi.org/10.1080/14786440109462720>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn, 2011: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Pospelov, G.S., 1988. Artificial Intelligence: the Basis for the New Information Technology.
- Preston F. W., and Henderson J.. 1964. *Fourier series characterization of cyclic sediments for stratigraphic correlation*. Kansas Geological Survey,
- Qing-You Zhang and João Aires-de-Sousa., 2007. Random Forest Prediction of Mutagenicity from Empirical Physicochemical Descriptors, *J. Chem. Inf. Model.* 2007, 47, 1-8.
- Ramadevi, R., Prakash, V., Sheela Rani, B., 2012. Role of hidden neurons in an Elman recurrent neural network in classification of cavitation signals. *Int. J. Comput.* 37 (7), 9–13 2012
- Reddy, R., and Bonham-Carter, G., 1991. A decision-tree approach to mineral potential mapping in snow lake area, manitoba. *Canadian Journal of Remote Sensing*, 17(2):191–200, 1991. doi: 10.1080/07038992.1991.10855292.
- Rezaee, M. R., Lemon, N. M., & Seggie, R. J., 1997. Tectonic fingerprints in siderite cement, Tirrawarra Sandstone, southern Cooper Basin, Australia. *Geological Magazine*, 134(1), 99-112.
- Romain , B., 2003. Modèles Impulsionnels de Réseaux de Neurones Biologiques Thèse Doctorat , Paris 2003 pp 03-10.
- Rosenblatt, F., 1962. Principales of Neurodynamics NEW YORK : Spartan.
- Rossi, C., Kalin, O., Arribas, J., Tortosa, A., 2002. Diagenesis, provenance and reservoir quality of Triassic TAG sandstones from Ourhoud field, Berkine (Ghadames) Basin, Algeria . *Marine and Petroleum Geology* 19 (2002) 117-145.
- Saad, M. Benrejeb., 2006. Optimisation multicritère par Pareto-optimalité de problèmes d'ordonnancement tenant compte du coût de production », *Revue Sciences et Technologies de l'Automatique*, e-STA, 2006.
- Saadi K., 2009. Contribution à l'étude de la Relation structure chimique-odeur Utilisation de la technique Random Forest " Application à la famille des pyrazines " mémoire de magister, université kasdi merbah ouargla. P-42.
- Salehi, M. M., Rahmati, M., Karimnezhad, M., & Omidvar, P., 2017. Estimation of the non records logs from existing logs using artificial neural networks. *Egyptian Journal of Petroleum*, 26(4), 957-968.
- Saporta, G., 1995. probabilités, analyse des données et statistiques. Technip , paris, 1990.
- Schütt , K. T., F. Arbabzadah, S. Chmiela, K. R. Müller, and A. Tkatchenko. Quantum-chemical insights from deep tensor neural networks. *Nat. Commun.*, 8:13890, Jan. 2017. ISSN 2041-1723. doi: 10.1038/ncomms13890. URL <http://dx.doi.org/10.1038/ncomms13890>.
- Schwartz, L. M., & Banavar, J. R., 1989. Transport properties of disordered continuum systems. *Physical Review B*, 39(16), 11965

- Schwarzacher, W., 1972. The semi-markov process as a general sedimentation model. In *Mathematical Models of Sedimentary Processes*, pages 247–268. Springer, 1972.
- Scott, A.j., Wheller, D.A., 1999. Regional trends in Triassia fluvial-lacustrine cycle, Berkine Basin, Algeria, American association of petroleum of petroleum Geologists Annual Meeting Expanded Abstracts, San Antonio, Texas, United States, p. 1126.
- Sebtosheikh, M.A., Motafakkerfard, R., Riahi, M.I., and Moradi, S., 2015. Separating Well Log Data to Train Support Vector Machines for Lithology Prediction in a Heterogeneous Carbonate Reservoir Iranian Journal of Oil & Gas Science and Technology 4 (5), 01-14.
- Seeman U, SchererM., 1984. Volcanoclastics as potential hydrocarbon reservoirs, Clay Mineralogy, vol. 9 (pg. 457-70)
- Serra, O., 1984. The acquisition of logging data. Elsevier.
- Sheela, K.G., Deepa, S.N., 2013. Review on Methods to Fix Number of Hidden Neurons in Neural Networks Mathematical Problems in Engineering, vol. 2013 <https://doi.org/10.1155/2013/425740>. Article ID 425740, 11 pages, 2013.
- Shen, C.. 2018. A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resources Research*, 54(11):8558–8593,.
- Simon B., 2009. Forêts Aléatoires: De l'Analyse des Mécanismes de Fonctionnement à la Construction Dynamique. Apprentissage [cs.LG]. Université de Rouen, 2009. Français. tel-00598441
- Singh, S., Kanli, A. I., & Sevgen, S., 2016. A general approach for porosity estimation using artificial neural network method: a case study from Kansas gas field. *Studia Geophysica et Geodaetica*, 60(1), 130-140.
- Souadnia, S., & Mezghache, H., 2017. Modélisations géostatistiques du faciès petrophysique du réservoir d'hydrocarbure de TAGI-Hassi Berkine Sud Geostatistical modeling of petrophysical facies of hydrocarbon reservoir TAGI-Hassi Berkine South.
- Sun, X., Deng, J., Gong, Q., Wang, Q., Yang, L., Zhao, Z., 2009. Kohonen neural network and factor analysis based approach to geochemical data pattern recognition. *Journal of Geochemical Exploration* 103, 6-16. <https://doi.org/10.1016/j.gexplo.2009.04.002>
- Szabó, N. P., & Dobróka, M., 2017. Robust estimation of reservoir shaliness by iteratively reweighted factor analysis. *Geophysics*, 82(2), D69-D83.
- Szabó, NP., 2011. Shale volume estimation based on the factor analysis of well logging data. *Acta Geophysica* 59, 935-953. <https://doi.org/10.2478/s11600-011-0034-0>
- Tarek, A., 2004. Introductions au systèmes Adaptatifs à base d'apprentissage Laboratoire A2SI-ESIEE-Paris.
- Tartakovsky, D.M., 2004. "Delineation of geologic facies with statistical learning theory". In: *Geophys. Res. Lett.* 31.18, p. 121. issn: 0094-8276. doi: 10.1029/2004GL020864.: <http://doi.wiley.com/10.1029/2004GL020864> (cit. on p. 20).
- Thomas, D.C., Pugh, V.J., A., 1987. Statistical analysis of accuracy and reproducibility of standard core analysis. *Proceeding of Society of Core Analysis*, Paper no. 8701.
- Timur, A., 1968. An investigation of permeability, porosity, & residual water saturation relationships for sandstone reservoirs. *The Log Analyst*, 9(04).
- Touzet, C., 1992. Les Réseaux de Neurones Artificiels – Introduction au Connexionisme.
- Turner, P., Pilling, D., Walker, D., Binnie, J., Sabaou, N., 2001. Sequence stratigraphy and sedimentology of the late Triassic TAGI (Blocks 401 – 402. Berkine, Algeria), *Marine Petroleum Geology* 18, 959 – 981.
- Turner, P., Pilling, D., Walker, D., Exton, J., Binnie, J., & Sabaou, N., 2001. Sequence stratigraphy and sedimentology of the late Triassic TAG-I (Blocks 401/402, Berkine Basin, Algeria). *Marine and Petroleum Geology*, 18(9), 959-981.
- Vaferi, B., Eslamloueyan, R., Ayatollahi, S., 2011. Automatic recognition of oil reservoir models from well testing data by using multi-layer perceptron networks. *Journal of Petroleum Science and Engineering*, 77, 254-262.
- Vapnik., V., 1995. *The nature of statistical learning*“, theory.spring-Verlag, New York,USA,1995
- W.E.C. 2007. Well Evaluation Conference.
- Walsh, J. B., & Brace, W. F., 1984. The effect of pressure on porosity and the transport properties of rock. *Journal of Geophysical Research: Solid Earth*, 89(B11), 9425-9431.

- Wang, H., J. F. Wellmann, Z. Li, X. Wang, and R. Y. Liang., 2017a. A segmentation approach for stochastic geological modeling using hidden markov random fields. *Math. Geosci.*, 49(2):145–177, Feb. 2017a. ISSN 1874-8961, 1874-8953. doi: 10.1007/s11004-016-9663-9. URL <https://doi.org/10.1007/s11004-016-9663-9>.
- Wendt, W.A., Sakurai, S., Nelson, P.H., 1985. Permeability prediction from well logs using multiple regression, in: L.W. Lake, H.B.J. Carroll (Eds.), *Reservoir Characterization*, Academic Press, New York.
- Whitley, D., 1994. A genetic algorithm tutorial. *Statistics and computing*, 4(2), 65-85.
- Willmott CJ, 1981. On the validation of model. *Phys Geogr* 2:184–194.
- Wood, D.A., 2019. Predicting porosity, permeability and water saturation applying an optimized nearest-neighbour, machine-learning and data-mining network of well-log data, *Journal of Petroleum Science and Engineering*(2019), doi: <https://doi.org/10.1016/j.petrol.2019.106587>.
- Xue, G., Datta-Gupta, A., Valko, P., Balsingame, T., 1996. Optimal transformations for multiple regression: application to permeability estimation from well logs. SPE 35412 Presented at the Improved Oil Recovery Symposium, Tulsa, Ok, U.S.A. 21 April, 1996, 9 pages. <https://doi.org/10.2118/35412-PA>
- Yahi, N., 1999. Petroleum generation and migration in the Berkine (Ghadames) basin, eastern Algeria: an organic geochemical and basin modelling study (No. PreJuSER-39900). *Erdöl und Geochemie*.
- Yongchao Xue, Linsong Cheng, Jianye Mou, Wenqi Zhao, 2014. A new fracture prediction method by combining genetic algorithm with neural network in low-permeability reservoirs, *Journal of Petroleum Science and Engineering*, Volume 121, Pages 159-166, ISSN 0920-4105, <https://doi.org/10.1016/j.petrol.2014.06.033>.
- Yuste, R & Tank, D.W., 1996. Dendritic integration in mammalian neurons, a century after Cajal, *Neuron*, 16, pp. 701–16.
- Zargar, G., Bagheripour, P., Asoodeh., M., 2014. Fuzzy Modeling of Volume Reduction of Oil Due to Dissolved Gas Run off and Pressure Release. *Journal of Petroleum Exploration and Production Technology*, DOI: 10.1007/s13202-014-0099-9.
- Zerrouki Ahmed Ali, 2015. Contribution à la caractérisation des réservoirs fissurés du champ de Hassi Messaoud par classement flou, réseaux de neurones artificiels et magnétisme des roches. *Sciences de la Terre*. Université M'Hamed Bougara-Boumerdès, 2015.
- Zerrouki, A.A., Aïfa, T., Baddari, K., 2014. Prediction of natural fracture porosity from well log data by means of fuzzy ranking and an artificial neural network in Hassi Messaoud oil field, Algeria. *J. Petrol. Sci. Eng.* (2014).
- Zhang, J., Li, L., Ping, S., 2006. Neural Approach for Calculating Permeability of Porous Medium *Chinese Phys. Lett.* 23, 1009–1011
- Zhao, X., and Mendel, J. M., 1988. Minimum-variance deconvolution using artificial neural networks. *SEG Technical Program Expanded Abstracts*, 1988. <https://library.seg.org/doi/pdf/10.1190/1.1892433>
- Zuou, R., Xiong, Y., Wang, J., and Carranza, E. J. M.. 2019. Deep learning and its application in geochemical mapping. *Earth-science reviews*.

ANNEXES

Annexe 1: Rappel sur les méthodes statistiques utilisées

1. Analyse statistique monovariée

Ce traitement permet de déterminer les paramètres statistiques de la distribution de chaque variable, de positions centrales (moyenne, le mode,...etc.), et de dispersion (variance, écart type, coefficient de variation...etc.) des éléments chimiques. L'histogramme d'une variable donnée visualise la répartition des effectifs dans un certain intervalle de variation prise par cette variable. L'allure générale de l'histogramme et/ou de la courbe des fréquences cumulées permet d'apprécier l'homogénéité ou l'hétérogénéité des données; ainsi que la loi de distribution des variables.

2. Analyse bivariée

Le traitement statistique bivarié consiste à étudier deux variables aléatoires simultanément et de déterminer la relation qui existe entre elles, à l'aide, entre autre du calcul des coefficients de corrélation simples. Elles sont calculées pour chaque couple de variables, où celles-ci montrent clairement les relations mutuelles qui existent entre les variables.

Le coefficient de corrélation est calculé par la formule suivante :

$$r = \text{COV}(x, y) / s_x \cdot s_y$$

r : Coefficient de corrélation simple de la population ;

s_x et s_y : L'écart type des deux variables

$\text{COV}(x, y)$: la covariance de variables x et y .

Il est à rappeler que : $\text{cov}(x,y) = E(xy) - E(x) \cdot E(y)$

Les coefficients de corrélation simples sont regroupés dans un tableau récapitulatif appelé "**matrice de corrélation**". Le coefficient de corrélation varie entre -1 et +1.

La signification de ce dernier dépend du nombre d'individus utilisés pour son calcul. Une façon empirique de calculer la signification c'est de calculer la valeur de :

Où n est le nombre des données utilisées.

$$R_{\text{signif}} = \frac{2}{\sqrt{n-1}}$$

3. Analyse multivariée

L'analyse des données multivariée désigne l'ensemble des méthodes qui permettent de collecter, d'organiser, d'analyser et de présenter les données. Ces méthodes utilisent des techniques géométriques permettant de mettre en évidence des groupes d'individus

homogènes, des liens entre individus et variables, des ressemblances ou des différences entre variables et surtout de représenter les données initiales d'un Tableau R (n,p) par un nombre plus petit sans perdre l'information.

Soit un tableau (matrice) R(n,p) où n est le nombre de lignes et p le nombre de

Représentation géométrique

Les données peuvent être représentées graphiquement en nuage de points-lignes (Représentation R^p) et/ou en nuage de point variables (Représentation dans R^n) (Fig. 1).

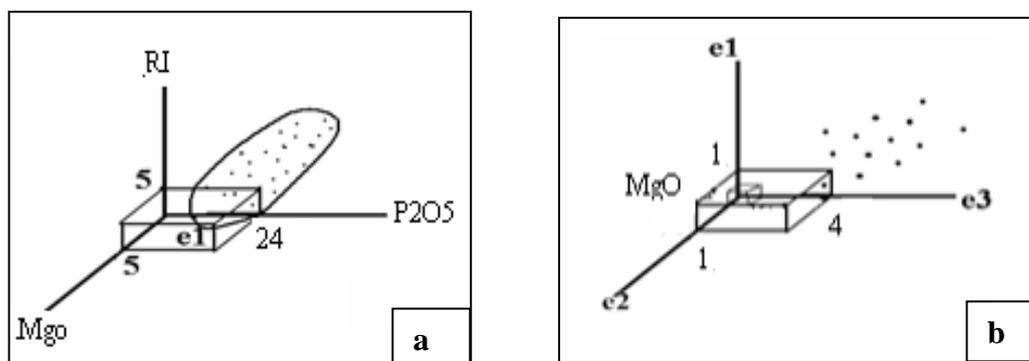


Fig. 1– Représentation géométrique des données.

a)- point-lignes R^p , b) point-variables R^n

Dans R^n , chaque point est défini par n coordonnées. Deux points-lignes apparaissent proche l'un de l'autre, si les deux lignes ont des mesures semblables des variables ou individus et éloignés l'un de l'autre si les mesures sont différentes.

En présence d'un nuage de points dans un espace à n dimensions, on cherchera à saisir et comprendre la dispersion des points c'est-à-dire la signification des différents allongements des nuages de points.

3.1. Analyse en Composantes Principales – ACP

L'analyse en Composantes Principales (ACP) est l'une des plus anciennes méthodes de l'analyse factorielle. Elle est utilisée pour l'analyse des données qui sont représentées sous formes de tableaux de n individus et p variables. Soit R (n,p), ce tableau de donnée de terme générale r_{ij} .

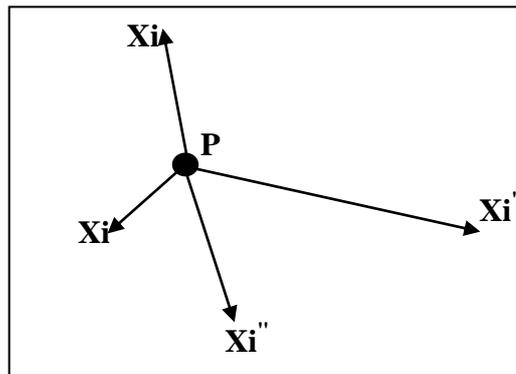
En ACP, ce qui important, c'est la forme du nuage des points par rapport aux axes (droites) qui passent par le centre de gravité du nuage. Il s'agit donc de rechercher les droites qui respectent au mieux l'ensemble des carrés des distances entre couples : C'est celles qui passent par le centre de gravité. D'où la nécessité de procéder à un changement de repère en centrant les variables de la matrice de données initiales. Si de plus les unités des variables sont

différentes alors il est préférables de les réduire pour supprimer l'effet unité c'est-à-dire échelle:

$$x_{ij} = \frac{r_{ij} - \bar{r}_j}{\sqrt{n} \cdot \sigma_j}$$

Ainsi après cette transformation (changement), la matrice d'inertie ne serait autre que la matrice de corrélation linéaire simple: $C = X^T X$

On appelle inertie du point N(I) par rapport au point P (Fig. 2), où l'expression suivant:



$$InP(I) = \sum_{i=1}^n m_i \|X_i - P\|^2$$

Fig. 2– Représentation géométrique d'inertie des points X_i ($i=1$ à n) par rapport au point

Après le calcul de C, il est nécessaire de calculer l'inertie du nuage de point dans les principaux axes d'allongement désignés par F_i ($i=1$ à N) (Fig. 109) en diagonalisant – c'est à dire en recherchant les premières valeurs propres (variances) et les vecteurs propres correspondants (Fig. 3).

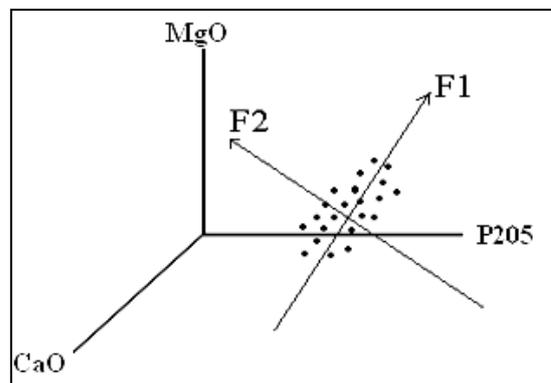


Fig. 3 – Représentation d'un nuage de points et les principaux axes d'allongement (Axes factoriels).

La diagonalisation de la matrice C (matrice d'inertie) permet d'obtenir les p vecteurs propres U_α et les p valeurs propres λ_α . Le vecteur U_α est connu par ces coordonnées dans l'espace R^p . Les coordonnées des points X^i sur les axes factorielles qui sont obtenus par:

$$F_\alpha(i) = X^i U_\alpha$$

$$F_\alpha(i) = \sum_{j=1}^k U_{\alpha j} X_{ij}$$

$F_\alpha(i)$ est le facteur de rang α associé à l'individu i : $F_\alpha(i)$ est la coordonnée de l'individu (échantillon) i sur l'axe factoriel U_α . Ce dernier a autant de composantes que de variables j . Dans la pratique l'inertie expliquée par l'axe factoriel:

$$T_\alpha = \frac{\lambda_\alpha}{In(I)} = \frac{\lambda_\alpha}{\sum_{\alpha=1}^p \lambda_\alpha}$$

Où $In(I)$ est l'inertie totale du nuage de points.

Généralement les résultats d'une ACP sont représentés sous forme graphique (Fig. 4).

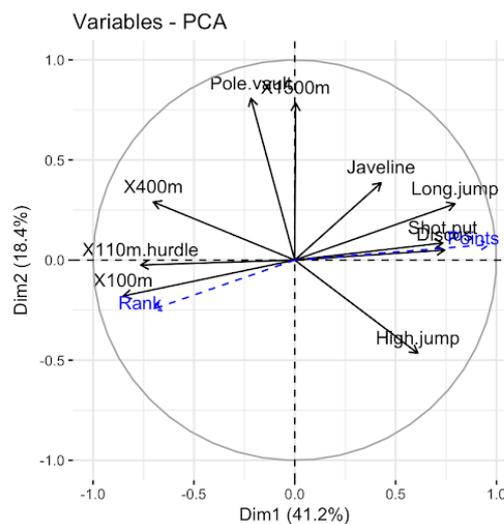


Fig. 4 – Représentation des facteurs de charge d'une ACP normée.